STOCHASTIC MODELS IN BIOLOGICAL SCIENCES
BANACH CENTER PUBLICATIONS, VOLUME 80
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WARSZAWA 2008

PREFACE

This volume contains papers presented at the workshop *Stochastic Models in Biological Sciences* held at the Stefan Banach International Mathematical Center in Warsaw, 29 May–2 June 2006. The workshop was organized in the framework of the European Science Foundation program "Phase Transitions and Fluctuation Phenomena for Random Dynamics in Spatially Extended Systems".

For several years now, biology has attracted the renewed interest of mathematicians and physicists. Apart from the development of new statistical algorithmic procedures to organize huge data sets, there is a need for analytical methods helping to understand physical and biochemical processes on the molecular level. The aim of our exploratory workshop on stochastic models in biology was to broaden the area of research of the mathematical statistical mechanics/stochastic processes community and to learn about issues and problems in biology which could be addressed by mathematics and theoretical physics. The workshop had a strong tutorial character. It attracted about 40 Ph.D. students and post-docs and introduced them to new interdisciplinary topics and research problems. All invited talks were pedagogical introductions to particular fields and such are the papers presented in this volume.

The leading motif of the workshop was the analysis of stochastic dynamics of finite populations. Stochastic effects can be modeled by Markov chains, stochastic processes or stochastic differential equations. Such stochastic models are analyzed from mathematical, physical and biological points of view. We identified five research areas which in our opinion are mature enough to allow an application of mathematical methods of stochastic analysis. More specifically, in many biological models we deal with systems of many interacting entities like genes in population genetics, species in evolutionary ecology, proteins in regulatory biochemical networks, or players in evolutionary games. The number of interacting objects is finite and often very small and therefore their stochastic fluctuations play a decisive role in the time evolution of a given system. Topics discussed during the workshop have included evolutionary genetics and ecology, molecular motors, genetic regulatory networks and evolutionary game theory.

Regulation of gene expression is a chemical process involving many coupled elementary chemical reactions. For this reason, it is frequently modeled by using classical chemical kinetics in which the system is represented by a set of differential equations describing the time evolution of molecular concentrations. However, gene expression processes occur in

very small volumes and involve very small numbers of molecules. Therefore, it is expected that significant random fluctuations occur.

Ofer Biham *et al.* present Monte-Carlo simulations of the master equation for certain regulatory systems. They show that in some cases, deterministic equations produce only steady state solutions but in the stochastic dynamics the system exhibits bistability and therefore functions as a switch.

Adam Bobrowski describes similar regulatory systems by piece-wise deterministic Markov processes—stochastically coupled ordinary differential equations. He discusses asymptotic (in time) behaviour of such processes and a limiting behaviour when certain parameters are large.

Long-run behaviour of different populations can be often described within game-theoretic models. In replicator dynamics, infinite populations evolve deterministically toward stationary points. However, we have to take into account random matching of players, random perturbations of their rationality in social models, and mutation and recombination in genetic models.

Jens Christian Claussen describes various stochastic dynamics of finite populations of interacting players such as the Moran process or a local update process. He then performs the infinite population limit of appropriate Fokker-Planck equations and obtains respectively a replicator dynamics or an adjusted replicator dynamics.

György Szabó discusses mechanisms of cooperation in evolutionary games of the Prisoner's Dilemma. He studies effects of pay-offs and noise on the level of cooperation in various stochastic dynamics of finite populations. In particular, he shows that in some models, the noise actually enhances the cooperative behaviour.

The construction of nonequilibrium statistical mechanics is expected to take great advantage from the study of biological models where nonequilibrium fluctuations are important. Most relevant studies are those of biological transport and its modeling via variations of exclusion models, and of molecular motors which are often seen as the small-scale chemical engines by which life processes are sustained. Physically inspired questions of efficiency and of the balances between dissipation and production can be studied here in a fully biological context.

The paper of Mauro Mobilia *et al.* discusses variations on nonequilibrium interacting particle systems as they are known in the literature on stochastic lattice gases. Here, however, their biological realization and relevance is stressed and it is sought to identify some generic principles of active transport.

The paper by Reinhard Lipowsky and Steffen Liepelt addresses molecular motors in the framework of models of nonequilibrium statistical mechanics. In particular, the thermodynamics of the motor systems and its ATP hydrolysis is discussed in the framework of recent developments in the understanding of entropy production.

Population genetics is concerned with the study of the genetic composition of populations. This composition may be changed by selection, mutation, recombination, mating structure, migration, and other genetic, ecological, and evolutionary factors. Major themes are the investigation of the mechanisms that generate and maintain genetic variability, the study of how this variation affects evolutionary change, and the inference

of past evolutionary events from contemporary DNA sequence data. Thus, population genetics provides the basis for understanding biological evolution.

Ellen Baake and Robert Bialowons study two stochastic population models with mutation and selection, the first relying on a multitype branching process, the second being the Moran model in which lineages are no longer independent. They analyze the properties of the stationary distributions forwards in time, which result from evolution, and the asymptotic distributions backwards in time. In both backwards models, the authors concentrate on the ancestry of single individuals rather than the genealogies of samples of individuals.

Michael Baake investigates the evolution of genetic sequences with repeated units, such as microsatellites. The copy numbers of these units vary within and between species. Baake analyzes a class of recombination models with single crossovers and studies the resulting equilibrium distributions. The nonlinear and infinite-dimensional nature of these models makes them interesting from a purely mathematical point of view: they can be viewed as nonlinear analogues of Markov chains.

Alison Etheridge's contribution also focuses on the interplay between forwards-in-time and backwards-in-time models. She starts by describing Kingman's, now classical, coalescent process which provides an elegant description of the genealogical relationship among individuals under neutral evolution. Then, she treats various important generalizations which include selection, spatial structure, or large-scale extinction-recolonization events.

Paul Higgs considers the way in which population genetics and phylogenetics are linked. In phylogenetics, species are typically represented by a single DNA sequence and properties of the substitution process are used to infer evolutionary (species) trees. This is clearly an over-simplification because species are genetically variable. The author demonstrates how phylogenetic modeling can benefit by taking population-genetic considerations into account.

Reinhard Bürger
Christian Maes
Jacek Miękisz

STOCHASTIC MODELS IN BIOLOGICAL SCIENCES
BANACH CENTER PUBLICATIONS, VOLUME 80
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WARSZAWA 2008

SCIENTIFIC COMMITTEE

Ellen Baake (Bielefeld University)
Anton Bovier (WIAS, Berlin)
Reinhard Bürger (University of Vienna)
Marek Kimmel (Rice University, Houston)
Roman Kotecký (Charles University, Prague)
Antti Kupiainen (University of Helsinki)
Christian Maes (Katholieke Universiteit Leuven)
Jacek Miękisz (University of Warsaw) — chair
Piotr Pokarowski (University of Warsaw)

LOCAL ORGANIZERS

Jacek Miękisz
Institute of Applied Mathematics
University of Warsaw

Michał Komorowski (University of Warsaw)
Aneta Miękisz (University of Warsaw)
Piotr Pokarowski (University of Warsaw)
Bartek Wilczyński (Polish Academy of Sciences)

INVITED SPEAKERS

Molecular motors:

- Erwin Frey, Ludwig-Maximilians-Universität München
- Anatoly Kolomeisky, Rice University
- Reinhard Lipowsky, Max Planck Institute of Colloids and Interfaces, Potsdam

Genetic regulatory networks:

- Ofer Biham, Hebrew University
- Andrzej Kierzek, University of Surrey
- Marek Kimmel, Rice University

Evolutionary game theory:
- Jens Christian Claussen, Christian-Albrechts-Universität Kiel
- Ulf Dieckmann, International Institute for Applied Systems Analysis, Laxenburg
- Josef Hofbauer, University College London
- György Szabó, Research Institute for Technical Physics and Material Science, Budapest

Evolutionary ecology:
- Alison Etheridge, University of Oxford
- Simon Levin, Princeton University
- Mirosław Dudek, Zielona Góra University
- Ryszard Rudnicki, Polish Academy of Sciences

Population genetics:
- Ellen Baake, Bielefeld University
- Michael Baake, Bielefeld University
- Adam Bobrowski, Polish Academy of Sciences
- Paul Higgs, McMaster University
- John Wakeley, Harvard University

STOCHASTIC MODELS IN BIOLOGICAL SCIENCES
BANACH CENTER PUBLICATIONS, VOLUME 80
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WARSZAWA 2008

PARTICIPANTS

1. Krzysztof Argasiński, Department of Environmental Sciences, Jagiellonian University, Cracow, Poland
2. Ellen Baake, Department of Biomathematics and Theoretical Bioinformatics, Bielefeld University, Bielefeld, Germany
3. Michael Baake, Faculty of Mathematics, Bielefeld University, Bielefeld, Germany
4. Wojciech Bartoszek, Department of Mathematics, Gdańsk University of Technology, Gdańsk, Poland
5. Przemysław Biecek, Institute of Genetics and Microbiology, Wrocław University, Wrocław, Poland
6. Ofer Biham, Racah Institute of Physics, Hebrew University, Jerusalem, Israel
7. Matthias Birkner, Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany
8. Marcelina Borcz, Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, Toruń, Poland
9. Tomasz Borowik, Institute of Physics, Wrocław University of Technology, Wrocław, Poland
10. Katarzyna Bońkowska, Institute of Genetics and Microbiology, Wrocław University, Wrocław, Poland
11. Enrico Carlon, École Polytechnique Universitaire de Lille and Interdisciplinary Research Institute, Lille, France
12. Jens Christian Claussen, Theoretical Physics + Astrophysics, University of Kiel, Kiel, Germany
13. Ulf Dieckmann, International Institute for Applied Systems Analysis, Laxenburg, Austria
14. Andrzej Drzewiński, Department of Physics, Częstochowa University of Technology, Częstochowa, Poland
15. Mirosław Dudek, Institute of Physics, University of Zielona Góra, Zielona Góra, Poland
16. Alison Etheridge, Magdalen College, University of Oxford, Oxford, UK

17. Erwin Frey, Statistical and Biological Physics, Ludwig-Maximilians-Universität, Münich, Germany
18. Natalya Golovchak, Biological Faculty, Ivan Franko National University of Lviv, Lviv, Ukraine
19. Beata Hat, Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland
20. Paul Higgs, Department of Physics and Astronomy, McMaster University, Hamilton, Canada
21. Josef Hofbauer, Department of Mathematics, University College London, London, UK
22. Rafał Hrynyk, Institute of Physics, Wrocław University of Technology, Wrocław, Poland
23. Andrzej Kierzek, SBMS, University of Surrey, Guildford, UK
24. Marek Kimmel, Department of Statistics, Rice University, Houston, USA
25. Galina Kleveta, Biological Faculty, Ivan Franko National University of Lviv, Lviv, Ukraine
26. Rafał Kluszczyński, Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, Toruń, Poland
27. Anatoly Kolomeisky, Department of Chemistry, Rice University, Houston, USA
28. Michał Komorowski, Institute of Applied Mathematics, Warsaw University, Warsaw, Poland
29. Roman Kotecký, Department of Theoretical Physics, Charles University, Prague, Czech Republic
30. Simon Levin, Department of Ecology and Evolutionary Biology, Princeton University, Princeton, USA
31. Tomasz Lipniacki, Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland
32. Reinhard Lipowsky, Max Planck Institute of Colloids and Interfaces, Potsdam, Germany
33. Florian Lipsmeier, Faculty of Technology, Bielefeld University, Bielefeld, Germany
34. Christian Maes, Instituut voor Theoretische Fysica, Katholieke Universiteit Leuven, Leuven, Belgium
35. Anna Marciniak, Institute of Applied Mathematics, University of Heidelberg, Heidelberg, Germany
36. Jacek Miękisz, Institute of Applied Mathematics, University of Warsaw, Warsaw, Poland
37. Mauro Mobilia, Arnold Sommerfeld Center for Theoretical Physics, Ludwig-Maximilians-Universität, Münich, Germany

38. Robert Mroczkowski, Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, Toruń, Poland
39. Szymon Niewieczerzał, Institute of Physics, Polish Academy of Sciences, Warsaw, Poland
40. Joanna Olesiak, Department of Chemistry, Wrocław University of Technology, Wrocław, Poland
41. Grzegorz Paździor, Institute of Physics, Wrocław University of Technology, Wrocław, Poland
42. Filip Piękniewski, Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, Toruń, Poland
43. Piotr Pokarowski, Institute of Applied Mathematics, University of Warsaw, Warsaw, Poland
44. Wojciech Puławski, Institute of Physics, Polish Academy of Sciences, Warsaw, Poland
45. Tobias Reichenbach, Arnold Sommerfeld Center for Theoretical Physics, Ludwig-Maximilians-Universität, Münich, Germany
46. Szymon Rogoziński, Institute of Theoretical Physics, University of Silesia, Katowice, Poland
47. Ryszard Rudnicki, Institute of Mathematics, Polish Academy of Sciences, Katowice, Poland
48. Tatiana Ryabukha, Institute of Mathematics, Ukrainian National Academy of Sciences, Kiev, Ukraine
49. Leszek Rybicki, Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, Toruń, Poland
50. Viacheslav Shtyk, Institute of Mathematics, Ukrainian National Academy of Sciences, Kiev, Ukraine
51. Matthias Steinruecken, Faculty of Technology, Bielefeld University, Bielefeld, Germany
52. Erik van der Straeten, Department of Physics, Antwerp University, Antwerp, Belgium
53. Christina Surulescu, Institute of Applied Analysis and Numerical Simulation, University of Stuttgart, Stuttgart, Germany
54. Joanna Sułkowska, Institute of Physics, Polish Academy of Sciences, Warsaw, Poland
55. György Szabó, Research Institute for Technical Physics and Material Science, Budapest, Hungary
56. Janusz Szwabiński, Institute of Theoretical Physics, Wrocław University, Wrocław, Poland

57. Antonina Tarnovska, Biological Faculty, Ivan Franko National University of Lviv, Lviv, Ukraine
58. Angelo Valleriani, Max Planck Institute of Colloids and Interfaces, Potsdam, Germany
59. John Wakeley, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, USA
60. Radek Wieczorek, Institute of Mathematics, Polish Academy of Sciences, Katowice, Poland
61. Bartek Wilczyński, Institute of Mathematics, Polish Academy of Sciences, Warsaw, Poland
62. Ryszard Wojnar, Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland
63. Paweł Wykrzykowski, Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, Toruń, Poland
64. Samuel Zajíček, Center for Theoretical Study, Prague, Czech Republic

STOCHASTIC MODELS IN BIOLOGICAL SCIENCES
BANACH CENTER PUBLICATIONS, VOLUME 80
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WARSZAWA 2008

DISCRETE STOCHASTIC PROCESSES, REPLICATOR AND FOKKER-PLANCK EQUATIONS OF COEVOLUTIONARY DYNAMICS IN FINITE AND INFINITE POPULATIONS

JENS CHRISTIAN CLAUSSEN

Institute of Theoretical Physics and Astrophysics, Christian-Albrecht University Kiel
Leibnizstr. 15, 24098 Kiel, Germany
E-mail: claussen@theo-physik.uni-kiel.de

Abstract. Finite-size fluctuations in coevolutionary dynamics arise in models of biological as well as of social and economic systems. This brief tutorial review surveys a systematic approach starting from a stochastic process discrete both in time and state. The limit $N \to \infty$ of an infinite population can be considered explicitly, generally leading to a replicator-type equation in zero order, and to a Fokker-Planck-type equation in first order in $1/\sqrt{N}$. Consequences and relations to some previous approaches are outlined.

1. Introduction. Evolution is a biological process ubiquituosly taking place, acting on on several temporal, spatial, and taxonomic scales. Biological organisms are coded by their genetic sequence, and their ability to survive, following Darwin, is largely determined by their genes. However, even the relationship between genotype, phenotype and a fitness (e.g., defined by the reproduction rate) by no means is simple, as it maps an enormously high-dimensional space (the state space of all possible genomes) to a much lower-dimensional phenotype space, and finally to a one-dimensional space of fitness values.

The extinction of species, as observed in the fossile record, follows a complicated behaviour [Eld72] including intermittency-like long periods of stasis, the so-called "punctuated equilibria". It is still at debate whether external or cosmic influences account for these, or whether a purely dynamic mechanism of the evolutionary and ecological dynamics can produce this stylized fact. In this direction, the Bak-Sneppen model [Bak93] is a pioneering minimal model for the extinction of species, which however is difficult to relate to the biological scenario. On the one hand, it provides an intermediate modeling level, where biological observations can be explained to some extent, but on the other hand,

2000 *Mathematics Subject Classification*: Primary 91A22; Secondary 60J20, 92D50, 92D15.
Key words and phrases: evolutionary game theory, finite populations, asymmetric conflicts.
The paper is in final form and no version of it will be published elsewhere.

 © Instytut Matematyczny PAN, 2008

a mathematical treatment is still possible [Ban05, Pis97]. On an individual-based level, the Tangled Nature model [Jen04] approaches closer to biology. While the dynamics at large will deserve some further decades of research, simplified situations of coevolutionary dynamics, restricting to a finite number of species can be studied by formal models of evolutionary processes, and can be solved analytically in many cases to allow for an exemplaric insight. In this brief tutorial review, coevolutionary dynamics in finite populations is formulated within the framework of evolutionary game theory, providing a convenient common mathematical framework for biological, social and economical evolutionary processes of strategies, or genetic types.

1.1. *Game theory: The strategy of conflict.* Game theory was brought into play by von Neumann and Morgenstern [Neu53] as a minimal model to explain and predict the behaviour of humans in strategic situations, military or economic, assuming fully rational behaviour. In game theory, more specifically in two-player normal form games, agents $1 \dots N$ "play" a strategy, out of a finite set of possible pure strategies. They interact with a partner also playing one of those stragegies, and both receive a (real valued) payoff according to a so-called payoff matrix. This is best illustrated with an example. The most paradigmatic conflict situation described in game theory is the Prisoner's Dilemma. It is defined by the payoff matrix

$$\pi_{ij} = \begin{pmatrix} 3 & 0 \\ 5 & 1 \end{pmatrix}. \tag{1}$$

This is to be read as follows. The game is defined between two players, who can adopt two possible strategies: to cooperate (C) or to defect (D). The players act in parallel and are not informed about the opponent's move. Each player P (row player) playing strategy i then receives the payoff π_{ij} when meeting a player using strategy j (column player, O) as opponent. Obviously, here the conflict situation is symmetric, therefore the payoff matrix of the opponent's payoffs is given by the transpose π_{ij}^{T}. Often both payoff matrices are combined to

$$(\pi_{ij}^{\mathrm{P}}, \pi_{ij}^{\mathrm{O}}) = \begin{pmatrix} 3,3 & 0,5 \\ 5,0 & 1,1 \end{pmatrix}. \tag{2}$$

In symmetric conflicts, thereby only redundant information is added. In asymmetric conflicts, for obvious reasons often called *bimatrix games*, the opponent's payoffs, in general, can be different. We will analyze such cases in Section 6.

In all cases above, the players play their respective strategy with probability one, i.e., they play pure strategies. A straightforward generalization is that players — in a random and uncorrelated manner — can use *mixed strategies*, or *strategy profiles* s_n, i.e., possess a strategy vector (with components $(s_n)_i$ summing up to one, thus are elements of a simplex S_n), and gain payoffs which are the corresponding linear combination

$$u_n(s_n, s_m) = \sum_{ij} (s_n)_i \pi_{ij} (s_m)_j \tag{3}$$

for playing against one opponent (profile) s_n, and

$$u_n(s_n, \{s_m\}) = \sum_{ijm} (s_n)_i \pi_{ij} (s_m)_j \tag{4}$$

against a set of opponents indexed by m. Now, can mixed strategies exist that are successfull against all hypothetical sets of opponents? This brings us to the central concept of a Nash equilibrium [Nas51]: A Nash equilibrium is a mixed strategy, in which no single agent can improve its payoff by solely changing its own strategy.

1.2. *Nash equilibria.* A *strict Nash equilibrium* (resp., *Nash equilibrium*) is defined as a strategy profile $s^* = (s_1^*, \dots s_N^*)$ for which each agent's strategy s_n is a best (resp., best or equal) response to the strategies of the other players s_{-n}^*, i.e.

$$\forall_n \forall_{s \in S_n, s \neq s_n^*} \quad u_n(s_n^*, s_{-n}^*) \geq u_n(s, s_{-n}^*) \tag{5}$$

where $>$ instead of $\geq$ must hold for a strict Nash equilibrium.
Hereby $s_{-n}^* := (s_1, \dots s_{n-1}, s_{n+1}, \dots s_N)$ is called a strategy profile of the co-players (formally a set of $N-1$ profiles, usually it is understood that playing against all co-players linearly sums the payoffs received playing with each of them; so one can define the co-profile $(s_{-n}^*)_k := (N-1)^{-1} \sum_{i \neq n} (s_i)_k$ as a linearly averaged profile of the co-players; or define $u_n(., s_{-n}^*)$ as the sum of payoffs against the profile of each co-player).

Nash's theorem [Nas51] ensures that a normal-form game (as defined above) for a finite number of strategies and a finite number of players always posesses a Nash equilibrium. However, it can be degenerate or a mixed strategy (*mixed Nash equilibrium*). In the above Prisoner's Dilemma, $((1,0),(1,0))$, i.e. both players playing always "defect", is a Nash equlibrium; hence two memory-less agents playing it for one round have no incentive to cooperate. Conversely, the Stag-Hunt game $\pi_{ij} = \begin{pmatrix} 5 & 0 \\ 3 & 3 \end{pmatrix}$ has two symmetric Nash equilibria, $((1,0),(1,0))$ and $((0,1),(0,1))$, the Hawk-Dove game $\pi_{ij} = \begin{pmatrix} (V-C)/2 & V \\ 0 & V/2 \end{pmatrix}$ with $C > V$ has two nonsymmetric Nash equilibria, $((1,0),(0,1))$ and $((0,1),(1,0))$, the latter is an example of a population where both strategies are present.

1.3. *Evolutionary game theory and evolutionarily stable strategies.* Decades after game theory was invented, Maynard Smith and Price [May73] were the first to utilize its approaches to the understanding of biological conflict situations among whose the emergence of cooperation [Axe84] among animals and humans still is a continuously active field. In the dynamical picture of evolutionary game theory, the concept of the Nash equilibrium has its counterpart in the *evolutionarily stable strategies* (ESS), which are defined as a population in which a single mutant (changing to any of the possible strategies or genotypes) cannot invade the population. In infinite populations, traditionally described within the framework of replicator equations [Tay78, Hof79, Zee80], ESS appear as stable fixed points (see [Hof98] for a systematic treatment). In finite populations, however, this concept has to be refined [Now04, Tay04, Nei04, Wil04, Tra06b]. For a more detailed introduction into the field of evolutionary game theory, and its recent development, see the classical textbooks [Hof84, Hof98] and recent reviews [Sza07, Mie07].

2. Microscopic interactions: Game theory based on particle collision models. The description of social agent behaviour as interaction, or collision, of particles has been studied by Helbing establishing a quite general framework [Hel92a, Hel92b, Hel96],

which at that time had not further been taken up within evolutionary game theory or evolutionary dynamics.

One main limiting assumption of the mean-field type description discussed in the next sections is that any spatial organization can be neglected in a first order approximation. This approximation, as well as neglecting age structure and time delay [Alb04], however is not warranted in general.

2.1. *Patchy ecosystems.* To describe the full dynamics of ecological and evolutionary processes, it can be necessary to split up the population into parts, or patches, and to investigate an intermediate level of subdivided populations [Che03] or metapopulation dynamics [Hår02]. In general, a rich variety of dynamical scenarios can emerge; thus simplified models on lattices and graphs have been investigated widely.

2.2. *Spatial models.* Life typically is organized, to a very rough approximation, as a covering of the surface of earth, in competition for sunlight, solid ground, or hunting territory. So it is natural to investigate evolutionary dynamics of individuals located in a two-dimensional space [Now92], where collective phenomena can emerge [Her94]. Annother classical study by Lindgren and Nordahl [Lin94] investigated the spatial Prisoner's Dilemma game with strategies of different memory lengths. Spatial structure has been identified as one possible mechanism to promote cooperation [Now92]. In general, spatial game theory is capable of rich dynamical behaviour [Sza99, Szo04, Sza05, Tra04], as coarsening, segregation, and spiral waves. Likewise, models for opinion dynamics have been studied, as the Sznajd model [Szn00].

2.3. *Evolutionary dynamics on graphs.* The systematic understanding of evolutionary dynamics on graphs is a still developing field. An early investigation of coevolutionary dynamics (i.e., including frequency-dependence) on graphs has been given by Ebel and Bornholdt [Ebe02], investigating the iterated Prisoner's Dilemma on networks. On the small-world architecture, Szabó et al. [Sza04] investigated the Rock-Paper-Scissors game dynamics. In [Lie05, Oht06], fixation properties of evolutionary dynamics on graphs are studied, and special subgraphs have been identified to enhance or suppress fixation. For two strategies and non frequency-dependent fitness, Antal, Redner and Sood [Ant06] have provided exact results for fixation for the case of degree-uncorrelated graphs. Despite this significant progress [Sza07], a general theory of coevolutionary dynamics on graphs remains a formidable challenge.

2.4. *Unstructured population dynamics: Meanfield approach, or Pólya urn models.* The approximation of an unstructured population implies that individuals are undistinguishable (apart from their strategies or genotype), and individuals are chosen randomly from the population, for death, reproduction, and competition. Hence, the population can be viewed as a Pólya urn from which individuals are drawn for the stochastic process. All evolutionary processes discussed for finite populations in the remainder belong to this class of processes.

3. Microscopic evolutionary processes. Evolutionary processes have been widely considered in population genetics. For processes in discrete time, an important systematic

distinction has to be made between synchroneous or parallel [Baa97] update processes, as the Fisher-Wright process, and processes with overlapping generations, where one individual is replaced in each evolutionary step (see Fig. 1).

3.1. *Fisher-Wright process.* The process defined by Fisher [Fis30] and Wright [Wri31] in its original form is not frequency-dependent; the fitness of the individuals does not depend on the state of the population (given by the number of individuals in each of the strategies). The Fisher-Wright process is defined as follows. In each time step, all individuals reproduce with probabilities proportional to their fitness, until the same population size N is reached. An important case is given by *neutral evolution*, where all individuals have identical fitness. Hence, the discrete stochastic process that describes the time evolution resembles a random genetic drift, and no Darwinian principle is incorporated.

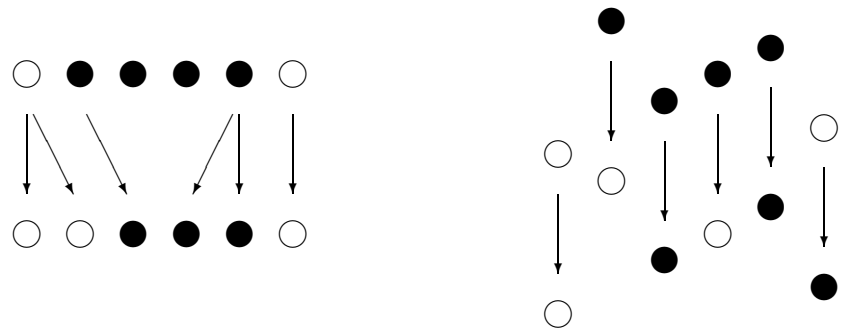


Fig. 1. Discrete stochastic evolution processes: Fisher-Wright process (left) as a typical synchroneous update process, and an overlapping generations process (right). In the Moran process and Local update the lifetime of the individuals however is nonidentical, due to the stochastic asynchroneous update. For the transition probabilities see text. The open and closed bullets can represent two different strategies (in a social system), or two different alleles at a specific genetic locus (in biology). The generalization to higher numbers of strategies or genetic types is straightforward.

Of course the Fisher-Wright process can be straightforwardly generalized to the case where the fitness of strategies depends on payoff gains obtained from game-theoretic interactions with other individuals. The finite population case has been considered recently by Taylor and Nowak [Tay06]. Here the transition matrix of the Markov process is by no means sparse; apart from the absorbing boundaries all matrix elements can be nonzero.

3.2. *Discrete stochastical processes for overlapping generations.* The evolutionary processes of Moran or Local update type provide us with a transition probability $T^{\pm}$ that the number of individuals i (being in the first strategy) increases or decreases by one, respectively.

For definiteness, we consider explicitly the Moran evolution dynamics in arbitrary 2×2 games [Cla05]. Given a finite population of N agents (two types/strategies A and B) interacting via in a game with the payoff matrix

$$P = \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \tag{6}$$

In the frequency-dependent Moran process, every agent interacts with a representative sample of the population, yielding the payoffs of A and B individuals as

$$\pi^A(i) = 1 - w + w\frac{a(i-1) + b(N-i)}{N-1}, \tag{7}$$

$$\pi^B(i) = 1 - w + w\frac{c\,i + d(N-1-i)}{N-1}, \tag{8}$$

i is the number of A individuals, and $1 - w$ is a background fitness.

3.3. *Moran process.* In its original form, also the Moran process [Mor62] is not frequency dependent. In this birth-death process, in each time step an individual is selected for reproduction, and subsequently a randomly selected individual dies. Here it is ensured that the total number of individuals N remains constant.

For two strategies, the Moran process therefore is a one-dimensional Markov process with a finite number of states from $i = 0, \ldots, N$. Taking up the Moran approach, Nowak, Taylor, Fudenberg and Sasaki [Now04, Tay04] investigated a frequency-dependent Moran process, defined as follows: In the frequency-dependent Moran process, selection for reproduction is proportional to the payoff compared to the average payoff, $\pi^A(i)/\langle\pi\rangle$. The probability per time step that a copy of an A agent is newborn then is $p^+ i/N$ (with p^+ as given in Table 1). It replaces a randomly chosen individual. Here, the fitness is evaluated after each individual took interaction according to a payoff matrix with all individuals in the population.

3.4. *Local update and imitation processes.* An apparent limitation of the Moran process with respect to biological situations is that individuals in each update step have to compete with the whole population (or, in real systems with a representative sample) of the population.

A more realistic setting is given by local or pairwise competetive interactions, which can be called imitation dynamics [Hof00] (in the concept of social strategies) or local update. These processes can be defined in a slightly different manner. In a non-symmetric definition of the (linear) local update process [Tra05], one individual is selected for reproduction, and the other for death. Then the strategy of the reproducing individual is changed (or kept) with probabilities that depend linearly on the difference of the payoffs of the two interacting individuals. Here, as is usually assumed in evolutionary game theory, the payoffs are evaluated by playing (on a shorter time scale) with the whole population, so these payoff differences still are evaluated in a global process.

3.5. *Nonlinear response: Local Fermi process.* A variant [Blu93, Sza98, Hau05] of the Local update process arises naturally when the game interaction payoff is one of many additive contributions to the reproductive fitness. In the case where the external contributions are large, and consist of many degrees of freedom that act like an external heatbath, the payoffs of the individuals appear as argument of Boltzmann factors, where a parameter w can be introduced as for the other processes ($w \to 0$ corresponds to weak selection), and here can be interpreted as a temperature. This pairwise comparison process has been studied in detail in [Tra06c, Tra07]. See Table 1 for the transition probabilities $T^{\pm}$ of the Moran process, Local update, and Fermi process.

Table 1. Comparison of three microscopic update processes and their respective differential equations for $x := i/N$ obtained in the deterministic limit of $N \to \infty$ (see Sec. 5.2).

Moran process	Local update	Fermi process
$p^+ = \dfrac{1-w+w\pi_i^A}{1-w+w\langle\pi_i\rangle}$	$p^+ = \dfrac{1}{2} + \dfrac{w}{2}\dfrac{\pi_i^A - \pi_i^B}{\Delta\pi_{\max}}$	$p^+ = \dfrac{1}{1+\mathrm{e}^{-w(\pi_i^A-\pi_i^B)}}$
$p^- = \dfrac{1-w+w\pi_i^B}{1-w+w\langle\pi_i\rangle}$	$p^- = \dfrac{1}{2} + \dfrac{w}{2}\dfrac{\pi_i^B - \pi_i^A}{\Delta\pi_{\max}}$	$p^- = \dfrac{1}{1+\mathrm{e}^{-w(\pi_i^B-\pi_i^A)}}$
$0 \le w \le 1$	$0 \le w \le 1$	$0 \le w \le \infty$
Transition probabilities for increase/decrease of i:		
$T^\pm = p^\pm \dfrac{i}{N}\dfrac{N-i}{N}$	$T^\pm = p^\pm \dfrac{i}{N}\dfrac{N-i}{N}$	$T^\pm = p^\pm \dfrac{i}{N}\dfrac{N-i}{N}$
In the $N \to \infty$ limit:		
$\dot{x} = \dfrac{wx(1-x)(\pi_x^A - \pi_x^B)}{1-w+w\langle\pi(x)\rangle}$	$\dot{x} = \dfrac{wx(1-x)(\pi_x^A - \pi_x^B)}{\Delta\pi_{\max}}$	$\dot{x} = x(1-x)\tanh(\dfrac{w}{2}(\pi_x^A - \pi_x^A))$
adjusted replicator eq.	*ordinary replicator eq.*	*replicator eq. after* [Tra06c]

4. Broadening of distributions in finite populations. For the Moran process, the strategy distribution is generated only by the inherent stochasticity of the finite population. At the borders, for $T_{0\to1}$ and $T_{N\to N-1}$ we assume a small mutation rate μ [Fud04]. While in the infinite population case the population density is peaked at the deterministic trajectory in the sense of a delta distribution, for a finite population it is broadened. Fortunately, the Moran process allows for an analytical treatment. The stationary distribution for an arbitrary payoff matrix can be expressed via Pochhammer symbols, rising factorials or gamma functions, for the general case of 2×2 games including nonvanishing background fitness [Cla05]. For illustration, let us consider the special cases of neutral evolution, constant fitness, an 'anticoordination game' and the Prisoner's dilemma,

$$P_{\mathrm{n}} = \begin{pmatrix} a & a \\ a & a \end{pmatrix}, \quad P_{\mathrm{c}} = \begin{pmatrix} a & a \\ c & c \end{pmatrix}, \quad P_{\mathrm{AC}} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad P_{\mathrm{PD}} = \begin{pmatrix} 3 & 0 \\ 5 & 1 \end{pmatrix}.$$

Internal (mixed) vs. external (pure) Nash equilibrium. For frequency dependent fitness and $w > 0$, the game can have an internal Nash equilibrium or an equilibrium in one of the absorbing states. A simple example with an internal Nash equilibrium is the 'anticoordination' game with $w = 1$, $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. The transition matrix here is $T_{i\to i+1} = \frac{N-i}{2N}, T_{i\to i-1} = \frac{i}{2N}$, describing a random walk with a drift towards the deterministic fixed point $i = N/2$. In equilibrium, we have $P_i\, T_{i\to i+1} = P_{i+1}\, T_{i+1\to i}$ for every i, thus

$$P_{i+1} = P_0 \prod_{j=0}^{i} \frac{N-j}{j+1} = P_0 \begin{pmatrix} N \\ i+1 \end{pmatrix}. \tag{9}$$

P_i is a binomial distribution around the equilibrium of the replicator dynamics.

The *Prisoner's Dilemma* has a Nash equilibrium for mutual defection, i.e. $i = 0$. At the borders, for $T_{0\to 1}$ and $T_{N\to N-1}$ we assume a small mutation rate μ [Fud04]. As $b = 0$, also state $i = 1$ is absorbing for $w = 1$ (two cooperators are needed to promote cooperation). Thus a small mutation rate μ has to be assumed also for $T_{1\to 2}$. Alternatively one could assume $w < 1$.

The four cases are summarized in Table 2. For *neutral evolution* the decay is very slow. With *constant fitness*, the probability decays approximately exponentially. For the

Table 2. The four payoff matrix cases considered in Fig. 2 under the assumption of a small mutation rate (see text). Without the frequency-dependence introduced by a game, the cases of constant fitness and neutral evolution known from mathematical genetics are recovered; the distribution keeps localized (for low mutation rates, mutants stay rare). For the analytical expressions of the distributions see [Cla05]. The distributions follow approximately a stretched exponential with the fit exponents γ as given in the table.

Payoff matrix	Distribution $\approx \exp(-bx^\gamma)$	Nash equilibrium (NE)?
$a = b = c = d$ Neutral evolution	$P_i \sim \frac{1}{i(N-i)}$	drift $\to i = 0$ and $\to i = N$
$a = b < c = d$ constant fitness	$\simeq$ exponential $\gamma = 0.87$	drift $\to i = 0$
$P_{AC} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$	$\simeq$ binomial $\gamma = 2.07$	$i = N/2$ "internal NE"
$P_{PD} = \begin{pmatrix} 3 & 0 \\ 5 & 1 \end{pmatrix}$	$\gamma = 0.63$	$i = 0$ "external NE"

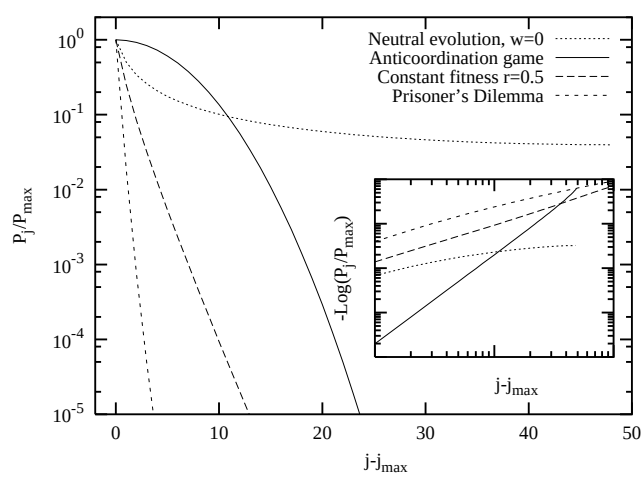


Fig. 2. Reproduced from [Cla05]. The invariant distributions P_j share different decay tails as a function of the distance $j - j_{\max}$ from the Nash equilibrium $j_{\max}$. Here $r = 1 - w + wa$. In the four cases considered here (shown for $N = 100$), the distributions can be approximated by stretched exponentials (which would appear as straight lines in appropriate scaling, the inset shows that the approximation is reasonable). For higher mutation rates see [Tra06a].

Prisoner's dilemma, the decay becomes slower with larger distance, while it becomes faster for the binomial distribution at the internal Nash equilibrium. The distributions are shown in Fig. 2. The decay of the distribution can be fitted by a $P \approx \exp(-bx^\gamma)$, $\gamma_{AC} = 2.06$, $\gamma_{CF} = 0.87$, $\gamma_{PD} = 0.63$. This corresponds to random motion in an anharmonic potential.

5. From microscopic equations to macroscopic equations Corresponding to the discreteness of states, we first describe the processes by a master equation, then formulate a Fokker-Planck (Kolmogorov forward) equation for large N, and finally compare the deterministic equations resulting from the limit $N \to \infty$.

5.1. *Limit of large populations: From master equation to Fokker-Planck equation.* The equation of motion (or time evolution equation) for the stochastic process can be formulated in terms of the master equation

$$\begin{aligned} P^{\tau+1}(i) - P^\tau(i) = \quad & P^\tau(i-1)T^+(i-1) - P^\tau(i)T^-(i) \\ & + P^\tau(i+1)T^-(i+1) - P^\tau(i)T^+(i) \end{aligned} \tag{10}$$

for the probability distribution, i.e., $P^\tau(i)$ is the probability to be in state i at time τ. For $N \gg 1$ we can proceed via a Kramers-Moyal expansion, defining $x = i/N$ and $t = \tau/N$. Then a formal Taylor expansion of T and $\rho(x,t) = N\,P^\tau(i)$ yields, considering only the two leading terms of the Taylor expansion,

$$\frac{d}{dt}\rho(x,t) = -\frac{d}{dx}\left[a(x)\rho(x,t)\right] + \frac{1}{2}\frac{d^2}{dx^2}\left[b^2(x)\rho(x,t)\right], \tag{11}$$

$$a(x) = T^+(x) - T^-(x), \tag{12}$$

$$b(x) = \sqrt{\frac{1}{N}\left[T^+(x) + T^-(x)\right]}. \tag{13}$$

For large, but finite N, this equation has the form of a Fokker-Planck equation. This allows to generalize the diffusion approximation (see e.g. [Dro01]) to coevolution. Since the internal noise is not correlated in time as subsequent update steps are independent, according to the Itô calculus a corresponding Langevin equation reads

$$\dot{x} = a(x) + b(x)\xi \tag{14}$$

where ξ is uncorrelated Gaussian noise and $b(x) = 0$ for $x = 0$ and $x = 1$. We here see that the noise is multiplicative and frequency-dependent.

5.2. *Limit of infinite populations: Replicator equation and adjusted replicator equation.* The leading order term does not vanish and describes the deterministic drift term. Surprisingly, the replicator equation [Tay78, Hof79, Zee81]

$$\dot{x} = \frac{wx(1-x)(\pi_x^A - \pi_x^B)}{\Delta\pi_{\max}} \tag{15}$$

is obtained for the Local update process [Tra05], whereas the adjusted replicator equation

$$\dot{x} = \frac{wx(1-x)(\pi_x^A - \pi_x^B)}{1 - w + w\langle\pi(x)\rangle} \tag{16}$$

is the deterministic limit of the Moran process [Tra05]. For other processes, other differential equations may be obtained, see Table 1.

The difference between the two equations may be best illustrated for the Prisoner's Dilemma. Here the (ordinary) replicator dynamics reads $\dot{x} = -x(1-x^2)$, whereas the adjusted replicator dynamics reads $\dot{x} = \frac{-x(1-x^2)}{\frac{1-w}{w}+1+3x-x^2}$. In this case, the stability of fixed points is preserved, as the additional denominator can be absorbed into a dynamical rescaling of the time scale, commonly known as velocity transform. However, in asymmetric conflicts the change of time scale matters, as will be detailed in the next section.

6. Asymmetric and cyclic games. Cyclicity of evolutionary dynamics intuitively is contradictory to the oversimplified picture of an absolute *fitness landscape*: any species having reached the maximal fitness value within a population will outcompete all others, and no cyclicity or oscillation can emerge. However, in ecological competition, oscillations of populations are quite generic, as known since Lotka [Lot20] and Volterra [Vol26] (see also [Mur04]). Of special interest are cases where species cyclically outcompete each other, as in Dawkins' caricature of mating strategies (see Sections 6.1–6.2), or the children's game *Rock-Paper-Scissors*, where rock crushes scissors, paper covers rock, and scissors cuts paper. This situation has been spotted in the territorial behaviour of lizards [Sin96, Zam00], and also in E.coli bacteria in vitro [Ker02] and in vivo [Kir04]. The Fokker-Planck equation, in analogy to Sec. 5.1, for the Rock-Paper-Scissors game has recently been given in [Rei06], and generalizations to other cyclic evolutionary games are straightforward. It is generally claimed that such cyclic coevolution promotes biodiversity [Csa02, Cla08]. However, its relevance in a more general picture of "evolution at large" remains an issue still to be investigated.

6.1. *Asymmetric conflicts (bimatrix games).* As mentioned in the introduction, the payoff matrix for the opponent can be different from that earned by the first player. To set this scenario into work, it is usually required that the game is played between two disjunct populations (as below, female and male), or that the interaction process itself is asymmetric, e.g., that one player is "initiative" and gains payoffs different from those earned in the opponents' role. Many social and economic situations bear such asymmetries, but they are often too weak to be significantly extracted from data.

An illustrative example of biological mating behaviour has been given by Dawkins [Daw76]. Male and female each can occur with two genetic strategies, "fast" (male: philanderer, female: "fast") and "slow" (male: "faithful", female: "coy"). The payoff benefit of a child is assumed to be b for both parents, the total cost of raising an offspring is $(-2c)$, and the prolonged courtship that coy females insists on add a burden of a to both parents. Coy females and male philanderers produce nothing and gain nothing. The cost $(-2c)$ is covered by both parents, except for philanderer males and fast females, where the female has to growup its offspring alone. This translates into the payoff matrix (with usual parameter choices $a = 3, b = 15, c = 10$, see also Fig. 3)

$$(\pi_{ij}^{\mathrm{M}}, \pi_{ij}^{\mathrm{F}}) = \begin{pmatrix} (b-c-a, b-c-a) & (b-c, b-c) \\ (0,0) & (b, b-2c) \end{pmatrix} = \begin{pmatrix} (2,2) & (5,5) \\ (0,0) & (15,-5) \end{pmatrix}. \tag{17}$$

If the four conditions $a > 0, c > 0, b - c > 0, 2c - a - b > 0$ are fulfilled, the game is cyclic. These conditions correspond to the arrows in Fig. 3. The cyclicity of the game is

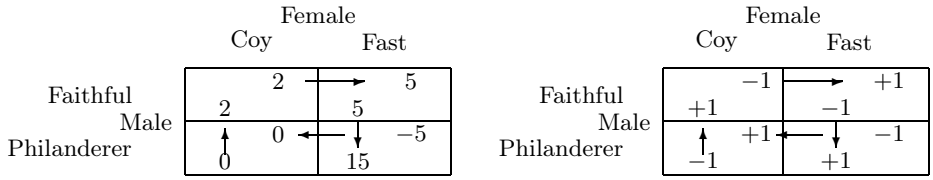


Fig. 3. Asymmetric payoff matrices for the Battle of the Sexes [Daw76]. Left: Original payoff matrix suggested by Dawkins. Right: Simplified payoff matrix, identical to the game "Matching pennies". In both cases, the left entry is for the row player, and the right entry (upon a common convention, shifted upwards, to enhance intuitive assignment) is for the column player. Arrows indicate the cyclic dominance.

preserved if we consider a normalized version (being equivalent to the payoff matrix of "Matching Pennies" played by two players):

$$(\pi_{ij}^{\mathrm{M}}, \pi_{ij}^{\mathrm{F}}) = \begin{pmatrix} (+1,-1) & (-1,+1) \\ (-1,+1) & (+1,-1) \end{pmatrix}. \tag{18}$$

For the cyclicity refer to Fig. 3. Let the cycle start in the lower right corner, where fast females and male philandrers are present. If males are philanderers, it pays for females to be coy (lower left corner); insisting on a long courtship period to make males invest more in the offspring (upper left corner). However, once most males are faithful, fast females are favored avoiding the costs of courtship (upper right corner). Subsequently, the male investment into the offspring is no longer justified, philanderers are again favored (lower right corner), and the cycle continues.

This game exhibits neutrally stable periodic orbits [May87] when described by the usual replicator equation approach [Tay78, Hof79, Zee81]. This would, according to Dawkins, lead to an infinitely lasting oscillation of strategies. As Dawkins argues, certain species (gibbon, stickleback, duck, fruit fly) have, in the course of real evolution, chosen to follow a fixed pair of strategies; for them the "Battle of the Sexes" has come to a rest. Of course this models relies on several assumptions, and in the remainder we will analyze how the conclusions change when considering different evolutionary processes and when explicitly considering the finiteness of the population.

6.2. *Counterintuitive behaviour for the "Battle of the Sexes".* The case of asymmetric conflicts opens the possibility of a conterintuitive finite-size effect: The sign of the average of $H = -xy(1-x)(1-y)$ (being a constant of motion for the ordinary replicator equation) changes for the Moran process above a critical population size, whereas the change remains positive (spiraling outwards) for the Local update and Fermi process in all finite populations (Fig. 4) [Tra05, Cla06, Cla07]. Hence in the $N \to \infty$ limit the Moran process shows a deterministic behaviour (that of the adjusted replicator equation), which is qualitatively different from that of the ordinary replicator equation. The reason is that the common velocity transformation, which can absorb the normalization denominator of the average fitness, here cannot be performed, as female and male population earn different payoffs and the denominators no longer coincide. The adjusted replicator equations

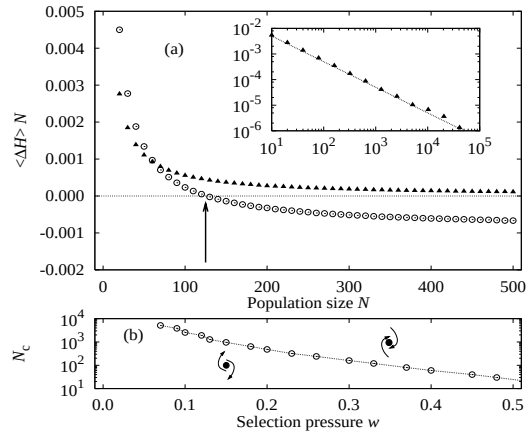


Fig. 4. Reproduced from [Tra05]. Drift reversal in the asymmetric game "Battle of the Sexes". Shown is the average change of H (which is a constant of motion in the ordinary replicator equation) for different population sizes. (a) For the Moran process (circles) above a critical population size the average change of H becomes negative so that trajectories spiral outwards on average. For the Local update, no drift reversal is observed and for $N \to \infty$ the change of H approaches zero (shown in the inset in double logarithmic plot). (b): Scaling of the critical population size with selection pressure w.

here read

$$\dot{x} = +2\frac{(x^2 - x)(2y - 1)}{\frac{1-w}{w} + (2y - 1)(2x - 1)}, \tag{19}$$

$$\dot{y} = -2\frac{(y^2 - y)(2x - 1)}{\frac{1-w}{w} - (2y - 1)(2x - 1)}. \tag{20}$$

Hence, if the two populations earn different average payoffs (here they differ by a factor -1), the fixed point stability of both types of replicator equations can differ.

References

[Alb04] J. Alboszta and J. Miękisz, *Stability of evolutionarily stable strategies in discrete replicator dynamics with time delay*, J. Theor. Biol. 231 (2004), 175–179.

[Ant06] T. Antal, S. Redner, and V. Sood, *Evolutionary Dynamics on Degree-Heterogeneous Graphs*, Phys. Rev. Lett. 96 (2006), 188104.

[Axe84] R. Axelrod, *The Evolution of Cooperation*, Basic Books, New York, 1984.

[Baa97] E. Baake, M. Baake, and H. Wagner, *Ising quantum chain is equivalent to a model of biological evolution*, Phys. Rev. Lett. 78 (1997), 559–562.

[Bak93] P. Bak and K. Sneppen, *Punctuated equilibrium and criticality in a simple model of evolution*, Phys. Rev. Lett. 71 (1993), 4083–4086.

[Ban05] C. Bandt, *The discrete evolution model of Bak and Sneppen is conjugate to the classical contact process*, J. Stat. Phys. 120 (2005), 685–693.

[Blu93] L. E. Blume, *The statistical mechanics of strategic interaction*, Games Econom. Behav. 5 (1993), 387–424.

[Che03] J. L. Cherry and J. Wakely, *A diffusion approximation for selection and drift in a subdivided population*, Genetics 163 (2003), 421–428.

[Cla05] J. C. Claussen and A. Traulsen, *Nongaussian fluctuations arising from finite populations: Exact results for the evolutionary Moran process*, Phys. Rev. E 71 (2005), 025101(R).

[Cla06] J. C. Claussen and A. Traulsen, *Fluctuations in coevolutionary dynamics and implications for multi-agent models*, in: D. Helbing (ed.), Proc. Potentials of Complexity Science for Business, Government, and the Media, Budapest, 2006, 411–419.

[Cla07] J. C. Claussen, *Drift reversal in asymmetric coevolutionary conflicts: Influence of microscopic processes and population size*, Eur. Phys. J. B 60 (2007), 391–399.

[Cla08] J. C. Claussen and A. Traulsen, *Cyclic dominance and biodiversity in well-mixed populations*, Phys. Rev. Lett. 100 (2008), 058104.

[Csa02] T. L. Czárán, R. F. Hoekstra and L. Pagie, *Chemical warfare between microbes promotes biodiversity*, Proc. Natl. Acad. Sci. USA 99 (2002), 786–790.

[Daw76] R. Dawkins, *The Selfish Gene*, Oxford University Press, New York, 1976.

[Dro01] B. Drossel, *Biological evolution and statistical physics*, Adv. Phys. 50 (2001), 209–295.

[Ebe02] H. Ebel and S. Bornholdt, *Coevolutionary games on networks*, Phys. Rev. E 66 (2002), 056118.

[Eld72] N. Eldredge and S. J. Gould, *Punctuated equilibria: an alternative to phyletic gradualism*, in: T. J. M. Schopf (ed.), Models in Paleobiology, Freeman Cooper, San Francisco, 1972, 82–115.

[Fis30] R. A. Fisher, *The Genetical Theory of Natural Selection*, Clarendon, Oxford, 1930.

[Fud04] D. Fudenberg and L. A. Imhof, *Imitation processes with small mutations*, Harvard Institute of Economic Research Discussion Paper No. 2050 (2004).

[Hår02] K. C. Harding and J. M. McNamara, *A unifying framework for metapopulation dynamics*, Amer. Naturalist 160 (2002), 173–185.

[Hau05] C. Hauert and G. Szabó, *Game theory and physics*, Am. J. Phys. 73 (2005), 405–414.

[Hel92a] D. Helbing, *A mathematical model for attitude formation by pair interactions*, Behavioral Science 37 (1992), 190–214.

[Hel92b] D. Helbing, Phys. A 181 (1992), 29-52; Phys. A 193 (1993), 241–258; Phys. A 196 (1993), 546–573.

[Hel96] D. Helbing, *A stochastic behavioral model and a "microscopic" foundation of evolutionary game theory*, Theory and Decision 40 (1996), 149-1-79.

[Her94] A. V. M. Herz, *Collective phenomena in spatially extended evolutionary games*, J. Theor. Biol. 169 (1994), 65–87.

[Hof79] J. Hofbauer, P. Schuster, and K. Sigmund, *A note on evolutionary stable strategies and game dynamics*, J. Theor. Biol. 81 (1979), 609–612.

[Hof84] J. Hofbauer and K. Sigmund, *Evolutionstheorie und dynamische Systeme*, Paul Parey, Berlin, 1984.

[Hof98] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics*, Cambridge University Press, Cambridge, 1998.

[Hof00] J. Hofbauer and K. H. Schlag, *Sophisticated imitation in cyclic games*, J. Evol. Econ. 10 (2000), 523–543.

[Jen04] H. J. Jensen, *Emergence of species and punctuated equilibrium in the Tangle Nature model of biological evolution*, Phys. A 340 (2004), 697–704.

[Ker02] B. Kerr, M. A. Riley, M. W. Feldman, and B. J. M. Bohannan, *Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors*, Nature 418 (2002), 171–174.

[Kir04] B. C. Kirkup and M. A. Riley, *Antibiotic-mediated antagonism leads to a bacterial game of rock-paper-scissors in vivo*, Nature 428 (2004), 412.

[Lie05] E. Lieberman, C. Hauert and M. A. Nowak, *Evolutionary dynamics on graphs*, Nature 433 (2005), 312–316.

[Lin94] K. Lindgren and M. Nordahl, *Evolutionary dynamics of spatial games*, Phys. D 75 (1994), 292.

[Lot20] A. J. Lotka, J. Am. Chem. Soc. 42 (1920), 1595–1599.

[May73] J. Maynard Smith and G. Price, *The logic of animal conflict*, Nature 246 (1973), 15.

[May87] J. Maynard Smith and J. Hofbauer, *The "battle of the sexes": A genetic model with limit cycle behavior*, Theoret. Population Biol. 32 (1987), 1–14.

[Mie07] J. Miękisz, *Evolutionary game theory and population dynamics* (2007), arXiv:q-bio/0703062, Lecture Notes in Math. (in press).

[Mor62] P. A. P. Moran, *The Statistical Processes of Evolutionary Theory*, Clarendon, Oxford, 1962.

[Mur04] J. D. Murray, *Mathematical Biology I*, Springer, 2004.

[Nas51] J. Nash, *Non-cooperative games*, Ann. Math. 54 (1951), 287–295.

[Nei04] D. B. Neill, *Evolutionary stability for large populations*, J. Theor. Biol. 227 (2004), 397–401.

[Neu53] J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, 1953.

[Now92] M. A. Nowak and R. M. May, *Evolutionary games and spatial chaos*, Nature 395 (1992), 826.

[Now04] M. A. Nowak, A. Sasaki, C. Taylor, and D. Fudenberg, *Emergence of cooperation and evolutionary stability in finite populations*, Nature 428 (2004), 646–650.

[Oht06] H. Ohtsuki, C. Hauert, E. Lieberman, and M. Nowak, *A simple rule for the evolution of cooperation on graphs*, Nature 441 (2006), 502–505.

[Pis97] Yu. M. Pis'mak, *Solution of the master equation for the Bak-Sneppen model of biological evolution in a finite ecosystem*, Phys. Rev. E 56 (1997), 1326–1329.

[Rei06] T. Reichenbach, M. Mobilia, and E. Frey, *Coexistence versus extinction in the stochastic cyclic Lotka-Volterra model*, Phys. Rev. E 74 (2006), 051907.

[Sin96] B. Sinervo and C. M. Lively, *The rock-paper-scissors game and the evolution of alternative male strategies*, Nature 380 (1996), 240–243.

[Sza98] G. Szabó and C. Tőke, *Evolutionary prisoner's dilemma game on a square lattice*, Phys. Rev. E 58 (1998), 69–73.

[Sza99] G. Szabó, M. A. Santos, and J. F. F. Mendes, *Vortex dynamics in a three-state model under cyclic dominance*, Phys. Rev. E 60 (1999), 3776–3780.

[Sza04] G. Szabó, A. Szolnoki, and R. Izsák, *Rock-scissors-paper game on regular small-world networks*, Journal of Physics A 37 (2004), 2599.

[Sza05] G. Szabó, J. Vukov and A. Szolnoki, *Phase diagrams for an evolutionary prisoner's dilemma game on two-dimensional lattices*, Phys. Rev. E 72 (2005), 047107.

[Sza07] G. Szabó and G. Fath, *Evolutionary games on graphs*, Phys. Rep. 446 (2007), 97–216.

[Szn00] K. Sznajd-Weron and J. Sznajd, *Opinion evolution in closed community*, Internat. J. Modern Phys. C 11 (2000), 1157–1165.

[Szo04] A. Szolnoki and G. Szabó, *Vertex dynamics during domain growth in three-state models*, Phys. Rev. E 70 (2004), 027101.

[Tay04] C. Taylor, D. Fudenberg, A. Sasaki and M. A. Nowak, *Evolutionary game dynamics in finite populations*, Bull. Math. Biol. 66 (2004), 1621–1644.

[Tay06] C. Taylor and M. A. Nowak, *Evolutionary game dynamics with non-uniform interaction rates*, Theoret. Population Biol. 69 (2006), 243–252.

[Tay78] P. D. Taylor and L. B. Jonker, *Evolutionary stable strategies and game dynamics*, Math. Biosci. 40 (1978), 145–156.

[Tra04] A. Traulsen and J. C. Claussen, *Similarity based cooperation and spatial segregation*, Phys. Rev. E 70 (2004), 046128.

[Tra05] A. Traulsen, J. C. Claussen, and C. Hauert, *Coevolutionary dynamics: from finite to infinite populations*, Phys. Rev. Lett. 95 (2005), 238701.

[Tra06a] A. Traulsen, J. C. Claussen, and C. Hauert, *Coevolutionary dynamics in large, but finite populations*, Phys. Rev. E 74 (2006), 011901.

[Tra06c] A. Traulsen, M. A. Nowak and J. Pacheco, *Stochastic dynamics of invasion and fixation*, Phys. Rev. E 74 (2006), 011909.

[Tra06b] A. Traulsen, J. M. Pacheco, and L. A. Imhof, *Stochasticity and evolutionary stability*, Phys. Rev. E 74 (2006), 021905.

[Tra07] A. Traulsen, M. A. Nowak and J. Pacheco, *Stochastic payoff evaluation increases the temperature of selection*, J. Theor. Biol. 244 (2007), 349–356.

[Vol26] V. Volterra, Mem. Accad. Lincei 2 (1926), 31; translation in: Animal Ecology, R. N. Chapman, McGraw Hill, 1931, 409–448.

[Wil04] G. Wild and P. D. Taylor, *Fitness and evolutionary stability in game theoretic models of finite populations*, Proc R. Soc. Lond. Ser. B Biol. Sci. 271 (2004), 2345–2349.

[Wri31] S. Wright, *Evolution in Mendelian populations*, Genetics 16 (1931), 97.

[Zam00] K. R. Zamudio and B. Sinervo, *Polygyny, mate-guarding, and posthumous fertilization as alternative male mating strategies*, Proc. Natl. Acad. Sci. USA 97 (2000), 14427.

[Zee80] E. C. Zeeman, *Population dynamics from game theory*, in: A. Nitecki, C. Robinson (eds.), Proceedings of an International Conference on Global Theory of Dynamical Systems, Lecture Notes in Math. 819, Springer, Berlin, 1980.

[Zee81] E. Zeeman, *Dynamics of the evolution of animal conflicts*, J. Theor. Biol. 89 (1981), 249–270.

STOCHASTIC MODELS IN BIOLOGICAL SCIENCES
BANACH CENTER PUBLICATIONS, VOLUME 80
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WARSZAWA 2008

ANCESTRAL PROCESSES WITH SELECTION: BRANCHING AND MORAN MODELS

ELLEN BAAKE

Technische Fakultät, Universität Bielefeld
Postfach 100131, 33501 Bielefeld, Germany
E-mail: ebaake@techfak.uni-bielefeld.de

ROBERT BIALOWONS

Institut für Mathematik und Informatik, Universität Greifswald
Jahnstr. 15a, 17487 Greifswald, Germany
E-mail: robert.bialowons@uni-greifswald.de

Abstract. We consider two versions of stochastic population models with mutation and selection. The first approach relies on a multitype branching process; here, individuals reproduce and change type (i.e., mutate) independently of each other, without restriction on population size. We analyse the equilibrium behaviour of this model, both in the forward and in the backward direction of time; the backward point of view emerges if the ancestry of individuals chosen randomly from the present population is traced back into the past.

The second approach is the Moran model with selection. Here, the population has constant size N. Individuals reproduce (at rates depending on their types), the offspring inherits the parent's type, and replaces a randomly chosen individual (to keep population size constant). Independently of the reproduction process, individuals can change type. As in the branching model, we consider the ancestral lines of single individuals chosen from the equilibrium population. We use analytical results of Fearnhead (2002) to determine the explicit properties, and parameter dependence, of the ancestral distribution of types, and its relationship with the stationary distribution in forward time.

1. Introduction. Like most dynamical processes, stochastic models of biological populations are usually considered in the forward direction of time. Ancestral processes arise if attention is shifted to the past history of a population, given its present state. This

2000 *Mathematics Subject Classification*: Primary 92D15; Secondary 60J70.

Key words and phrases: multitype branching, Moran model, mutation-selection model, ancestral selection graph.

The paper is in final form and no version of it will be published elsewhere.

 © Instytut Matematyczny PAN, 2008

backward point of view is particularly important in the area of *population genetics*, which is concerned with the genetic structure of populations under the joint action of evolutionary forces like mutation and selection. Here, the large time scales involved usually make it impossible to observe the dynamics over any relevant time span; one therefore resorts to taking samples from present-day populations, and inferring their history, in a probabilistic or statistical sense. More precisely, starting from *population sequence data* (sequences from a single locus analysed in a sample of individuals of the same species), three different sorts of inference questions can be addressed (see [26]):

1. What are the genetic parameters (mutation rates, selective intensities) governing the dynamics of the process?
2. The past history of the genetic types observed in the sample; for example, the age of a mutation (e.g., when has a disease gene arisen?)
3. The demographic history of the population from which the data was sampled; for example, when did the most recent common ancestor live?

For an overview of the area and the corresponding inference techniques, see [26]. Clearly, such endeavour relies crucially on a good knowledge of the population process viewed in the reverse direction of time.

We will discuss such ancestral processes here for two classes of models that describe the interaction of mutation and selection (or, more generally, type-dependent reproduction). We will first consider certain (multi-type) branching processes, where individuals mutate and reproduce independently of each other, and, in particular, independently of the current population size, which is, therefore, free to fluctuate. At the other extreme, we will consider the so-called Moran model, which assumes that birth events are strictly coupled to death events (any newborn individual replaces one of the individuals already present), so that the population size remains constant throughout.

The purpose of the article is twofold. The larger part of the paper will explain, in a tutorial manner, the forward and the ancestral processes in both types of models, with special emphasis on the asymptotic behaviour. We will start with the branching process (Sec. 2–5), because it is simpler due to the independence of individual lines. The corresponding backward process is analogous to the usual time reversal of a Markov chain, and directly leads to the composition of that part of the population that will become ancestral to future generations. Genealogies (of individuals sampled from today's population) are not relevant, and not required, in this context.

In constrast, reversing time in the Moran model is more intricate, but also yields more information. Genealogies of individuals sampled today play a central role here. As long as selection is absent (i.e., all individuals reproduce at the same rate), genealogies are easily constructed through the famous Kingman's coalescent [20, 21], which starts with a sample from today's population and proceeds backward, joining pairs of lines at certain rates and thus producing genealogical trees. This process is quite tractable due to the crucial fact that the lineages of the sampled individuals may be considered independently of the remaining population. The properties of the resulting genealogies are, therefore, quite well understood.

This lucky situation changes when individuals reproduce at different rates. Due to the coupling of birth and death events, individual lineages now cease to be independent of the (type composition of) the remaining population, and a more intricate construction must be used, which is known as the *ancestral selection graph*. This is no longer a tree, but a branching/coalescing graph, into which all possible genealogies are embedded, and the 'true' one (in the porbabilistic sense) can be extracted according to certain rules. We will put forward the Moran model, and its backward constructions, in Secs. 6–12.

So far, our exposition will mainly be a review of results available in the literature, but not easily accessible for the uninitiated. In Secs. 13 and 14, we will then present some new results on a two-type example with asymmetric mutation, which illustrates the presented concepts, and establishes connections between the Moran and the corresponding branching models.

2. Multitype branching: the mutation-reproduction model. Let S be a finite set of types (with $|S| > 1$), and consider a population of individuals, each characterised by a type $i \in S$. We will think of types as *genotypes*, since children inherit their parents' type. (Genetically speaking, our individuals are haploid, i.e., carry only one copy of the genetic information per cell, and types are alleles.)

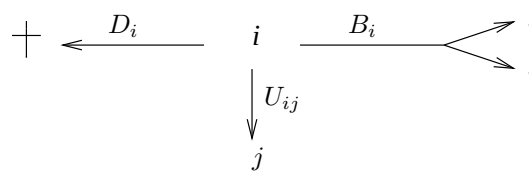


Fig. 1. The parallel mutation-reproduction model

We consider the most basic mutation-reproduction model, in which mutation and reproduction occur in parallel. As depicted in Fig. 1, an individual of type $i \in S$ may, at any instant in continuous time, do either of three things: It may split, i.e., produce a copy of itself (this happens at birth rate $B_i \geq 0$), it may die (at rate $D_i \geq 0$), or it may mutate to type $j, j \neq i$ (at rate $U_{ij} \geq 0$). This defines a *multi-type Markov branching process in continuous time* (see [1, Ch. V.7] or [17, Ch. 8] for a general overview, and [3] for the context considered here). That is, an i-individual waits for an exponential time with parameter $A_i = B_i + D_i + \sum_{j:j\neq i} U_{ij}$, and then dies, splits or mutates to type $j \neq i$ with probabilities $B_i/A_i, D_i/A_i$, and U_{ij}/A_i, respectively. The number of individuals of type j at time t, $Z_j(t) \in \mathbb{Z}_{\geq 0} := \{0, 1, 2 \dots\}$, is a random variable; the collection $Z(t) = \big(Z_j(t)\big)_{j\in S}$ is a random vector. The corresponding expectation is described by the first-moment generator $\boldsymbol{A} = \boldsymbol{U} + \boldsymbol{R}$. Here, $\boldsymbol{U}$ is the Markov generator $\boldsymbol{U} = (U_{ij})_{i,j\in S}$, where the mutation rates U_{ij} for $j \neq i$ are complemented by $U_{ii} := -\sum_{j:j\neq i} U_{ij}$ for all $i \in S$. Further, $\boldsymbol{R} := \mathrm{diag}\{R_i \mid i \in S\}$, where $R_i := B_i - D_i$ is the net reproduction rate (or *Malthusian fitness*) of an i-individual. More precisely, we have $\mathbb{E}^i(Z_j(t)) = (e^{t\boldsymbol{A}})_{ij}$, where $\mathbb{E}^i(Z_j(t))$ is the expected number of j-individuals at time t in a population started by a single i-individual at time 0.

3. Properties of the branching model. We will assume throughout that $\boldsymbol{A}$ (or, equivalently, $\boldsymbol{U}$) is irreducible. Perron-Frobenius theory then tells us that $\boldsymbol{A}$ has a principal eigenvalue λ (namely a real eigenvalue exceeding the real parts of all other eigenvalues) and associated positive left and right eigenvectors π and h, which will be normalised so that $\langle \pi, \mathbb{1} \rangle = 1 = \langle \pi, h \rangle$, where $\mathbb{1} = (1)_{i \in S}$ denotes the vector with all coordinates equal to 1, and $\langle .\,, . \rangle$ is the scalar product. We will further assume that $\lambda > 0$, i.e., the branching process is *supercritical*. This implies that the population will, in expectation, grow in the long run; in individual realisations, it will survive with positive probability, and then grow to infinite size with probability one, see (2) below. The asymptotic properties of our model are, to a large extent, determined by λ, π, and h. The left eigenvector π holds the stationary composition of the population, in the sense that

$$\lim_{t \to \infty} \frac{Z(t)}{\|Z(t)\|_1} = \pi \quad \text{with probability one, conditionally on survival,} \tag{1}$$

where $\|Z(t)\|_1 := \sum_{j \in S} Z_j(t)$ is the total population size. This is due to the famous Kesten-Stigum theorem, see [19] for the discrete-time original, and [1, Thm. 2, p. 206] and [13, Thm. 2.1] for continuous-time versions. Furthermore, with $R = (R_i)_{i \in S}$,

$$\langle \pi, R \rangle = \lambda = \lim_{t \to \infty} \frac{1}{t} \log \|Z(t)\|_1 \tag{2}$$

is the asymptotic growth rate (or equilibrium mean fitness) of the population. Here the first equality follows from the identity $\lambda = \langle \pi \boldsymbol{A}, \mathbb{1} \rangle = \langle \pi, \boldsymbol{A}\mathbb{1} \rangle = \langle \pi, R \rangle$, and the second is from [13] and holds with probability one in the case of survival. Finally, the i-th coordinate h_i of the right eigenvector h measures the asymptotic mean offspring size of an i-individual, relative to the total size of the population:

$$h_i = \lim_{t \to \infty} \mathbb{E}^i\big(\|Z(t)\|_1\big) e^{-\lambda t}. \tag{3}$$

For more details concerning this quantity, see [14] and [13] (for the deterministic and stochastic pictures, respectively).

In the above, we have adopted the traditional view on branching processes, which is forward in time. It is less customary, but equally rewarding, to look at branching populations backward in time. To this end, consider picking individuals randomly (with equal weight) from the current population and tracing their lines of descent backward in time (see Fig. 2). If we pick an individual at time t and ask for the probability that the type of its ancestor is i at an earlier time $t - \tau$, the answer will be $\alpha_i = \pi_i h_i$ in the limit when first $t \to \infty$ and then $\tau \to \infty$. Thus the distribution $\alpha = (\alpha_i)_{i \in S}$ describes the *population average* of the ancestral types and is termed the *ancestral distribution*, see [13, Thm. 3.1] for details. In contrast and for easy distinction, we will sometimes denote the stationary distribution of the forward process, π, as the *present distribution*.

If we pick individuals from the population at a very late time (so that its composition is given by the stationary vector π), then the type process in the backward direction is the Markov chain with generator $\boldsymbol{G} = (G_{ij})_{i,j \in S}$, $G_{ij} = \pi_j (A_{ji} - \lambda \delta_{ij}) \pi_i^{-1}$, as first identified by Jagers [15, 16]. Unsurprisingly, its stationary distribution is the ancestral distribution, α.

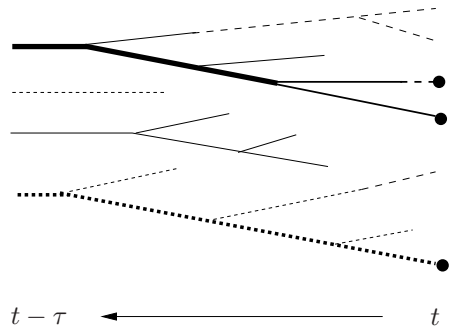


Fig. 2. The backward point of view. The various types are indicated by different line styles (solid, dashed and dotted). The fat lines mark the lines of descent defined by three individuals (bullets) picked from the branching population at time t. After coalescence of two such lines, the common ancestor receives twice the 'weight', as indicated by the extra fat line; this motivates the factor h_i in the ancestral distribution.

4. An example: Sequence space model with single-peaked landscape. Assume now that $S = \{0,1\}^L$, i.e., the type of an individual is characterised by a sequence of length L, of which every site may be either 0 or 1. (This may be interpreted as a sequence of matches (0) or mismatches (1) with respect to a reference sequence of nucleotides or amino acids, for example.) Assume that every site mutates at rate $\mu > 0$ from 0 to 1 or vice versa, independently of the other sites. Assume further that $D_i = 1$ for all $i = (i_1, i_2, \ldots, i_L) \in S$, $B_{00\ldots0} = 1 + s$, and $B_i = 1$ for $00\ldots0 \neq i$; i.e., the reference sequence has a selective advantage of $s > 0$ over all others, which are equally unfit. This is a branching process version of what is known as the *single-peaked landscape* of sequence evolution, which was introduced by Eigen [7] and has been discussed in numerous variants ever since (see [8] and [2] for reviews). Though not particularly realistic, it can serve as an instructive prototype model. Fig. 3 shows its stationary type distribution, π, as a function of the mutation rate. When there is little mutation, the population consists mainly of 'fit' individuals; with increasing μ, more and more mutants come up, until the favourable type is (nearly) lost, and the population turns into an (approximate) equidistribution on S, thus losing its genetic structure. This transition occurs in a fairly abrupt manner, around a value of $\mu \approx \mu_0 = s/L$, and becomes sharp (i.e., turns into a phase transition known as an *error threshold*) when $L \to \infty$. For a discussion of this and related threshold phenomena, see [14].

5. A two-type caricature. To capture the essence of the single-peaked model, let us condense it radically. Assume that there are only two types, i.e., $S = \{0,1\}$, with birth rates $B_0 = 1 + s$, $B_1 = 1$, death rates $D_0 = D_1 = 1$, and mutation rates $U_{01} = u\nu_1$ and $U_{10} = u\nu_0$; cf. Fig. 4. Here, $u > 0$ scales the overall mutation rate, whereas $\nu_0 > 0$ and $\nu_1 > 0$ ($\nu_0 + \nu_1 = 1$) determine the asymmetry of mutation. To mimic the single-peaked model, with the fit sequence corresponding to type 0 and all others collected into type 1, a simple choice is $\nu_0 = 1/L$ and $u = \mu L$; it approximates the fact that, in the sequence picture, μL is the total mutation rate away from $00\ldots0$, and μ is the maximal rate at

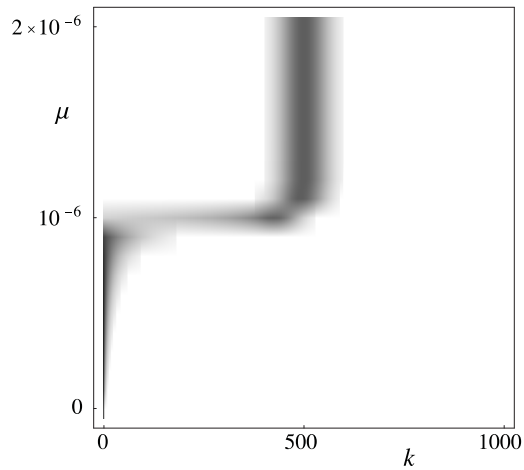


Fig. 3. The stationary distribution, π, for the single-peaked landscape with $L = 1000$ and $s = 0.001$, as a function of the mutation rate per site, μ. Types are lumped into classes: class k contains all sequences with $\#_1(i) = k$, where $\#_1(i)$ is the number of 1's in sequence i. Grey levels correspond to $\sum_{i:\#_1(i)=k} \pi_k$ (darker shading corresponds to larger values). The error threshold occurs at $\mu_0 \approx s/L = 10^{-6}$.

which an unfit individual can mutate back to the fit type. For this choice, π and α are shown in Fig. 10 as a function of u. Unsurprisingly, the stationary frequency of the fit type decreases (almost linearly) with u, until it gets close to zero (around $u_0 = s$). In contrast, the ancestral distribution consists almost exclusively of fit individuals until u gets close to u_0, and then declines to zero quickly. (For $\nu_0 \to 0$, the curves turn into piecewise linear (π) and piecewise constant (α).) Obviously, for $u < u_0$, the present distribution carries with it a 'comet tail' of unfit mutants, which are present at any time, but do not survive in the long run. They therefore rarely show up as ancestors, nearly all of whom are fit.

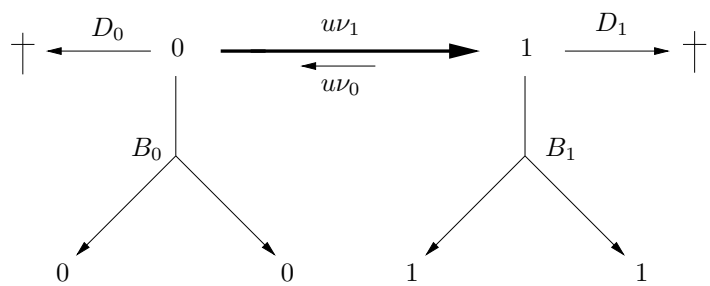


Fig. 4. The two-type mutation-reproduction process

6. The Moran model with selection. Branching processes have very nice features mathematically, because individuals reproduce independently of each other. Biologically, however, this independence can be quite unrealistic, because there is nothing to control population size. At the other extreme, one considers models for populations of *fixed* size.

The standard models in this context are the *Wright-Fisher model* and the *Moran model*. We will use the Moran model here because, in contrast to the Wright-Fisher model, it has a continuous-time version, which leads to the coalescent in a most immediate way; furthermore, it readily compares to our branching process. Luckily, however, the choice is not essential, since both the Wright-Fisher and the Moran models share the same diffusion limit (up to a factor of 2, see Remark 2 below); and it is only in this limit that the model can be tackled explicitly anyway, as will become clear below.

Let us now embark on the counterpart of the two-type branching model of the previous Section, namely, the Moran model with two types, mutation and selection.

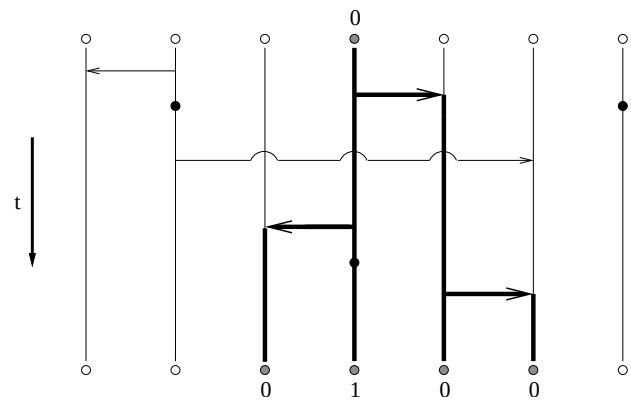


Fig. 5. The Moran model. The fat lines represent the descendants of the single type-0 individual marked grey at the top ($t = 0$). At the later time (bottom), its descendants consist of four individuals, one of whom has mutated to type 1 (mutation events are marked by blobs). Equivalently, the fat lines define the genealogy of the four 'grey' individuals at the bottom.

Moran models belong to the larger class of *interacting particle systems*, and are best explained in terms of their graphical representation, which, in our case, goes as follows (see Fig. 5). We have a fixed number of N individuals, each characterised by a type $i \in S = \{0, 1\}$, and represented by a vertical line; time runs from top to bottom. If an individual gives birth, the offspring inherits the parent's type and replaces an individual chosen randomly from the population (maybe its own parent), thus keeping population size constant. In the graphical representation, such events are displayed as arrows, with the parent at the base and the offspring at the tip. Type-0 individuals reproduce at rate $1 + s$, type-1 individuals at rate 1. This type dependence of reproduction rates leads to *selection*, which means that, on average, unfit individuals are replaced by fit ones more often than vice versa, thus leading to a net replacement of unfit individuals by fit ones. Independently of reproduction, individuals may mutate, at rate $u\nu_1$ from type 0 to type 1, and at rate $u\nu_0$ in the reverse direction. Equivalently, we can say that every individual mutates at rate u, and the ensuing type is i with probability ν_i, $i \in S$; one thus includes some 'empty events' that, in effect, do not change the current type. Mutation events are represented as blobs in the graphical representation.

The graphical representation is not meant to encode any spatial information – all events are independent of the location of the individuals. Therefore, all relevant information is contained in the process $\{Y_0(t), Y_1(t)\}_{t\geq 0}$, where $Y_i(t)$, $i \in S$, is the number of individuals of type i at time t (of course, $Y_0(t)+Y_1(t) = N$, but we retain both components for convenience here). If the current state is $(Y_0(t), Y_1(t)) = (k_0, k_1)$ $(k_0 = 0, \ldots, N)$, the possible transitions are

$$(k_0, k_1) \to \begin{cases} (k_0+1, k_1-1) \text{ at rate } u\nu_0 k_1 + (1+s)k_0 k_1/N \\ (k_0-1, k_1+1) \text{ at rate } u\nu_1 k_0 + k_0 k_1/N \end{cases}$$

(note that transitions to 'impossible' states ($k_i \notin \{0, \ldots, N\}$) occur at rate zero and are therefore excluded).

The standard strategy to make the process tractable is to consider the normalised process $\{X_0(t), X_1(t)\}_{t\geq 0} := \{Y_0(t), Y_1(t)\}_{t\geq 0}/N$, and take a diffusion limit, in which time is rescaled in units of N generations, and the limit $N \to \infty$ is taken so that $Ns \to \sigma$ and $Nu \to \theta$, where $\theta > 0$ and $\sigma \geq 0$. In this limit, the stationary distribution of type frequencies, in the forward direction of time, is well known: it is given by the density

$$f(x_0, x_1) = \frac{1}{C} x_0^{\theta\nu_0 - 1} x_1^{\theta\nu_1 - 1} e^{\sigma x_0}, \tag{4}$$

where C is the normalising constant. Eq. (4) is a classical result known as Wright's formula; see [11, Ch. 4,5] for a comprehensive review of diffusion theory in the context of population genetics.

In the diffusion limit, the above particle representation is no longer well-defined; a more advanced construction, known as the *look-down process* ([4, 5], see also [9, Ch. 5.2 and 5.3]) is required. In this tutorial introduction, we will only use the particle system for the finite population, and derive from it a good motivation for the backward construction in the diffusion limit.

Remark 1. In most of the literature, the Moran model with selection is defined in a slightly different version, in which mutation events can only occur on the occasion of reproduction (see, e.g., [6, Chap. 3.1]). But this *coupled* version has the same diffusion limit as our parallel model, provided the parameters are interpreted in a suitable way; actually, one effect of the diffusion limit is to decouple mutation and selection, which are both rare events in the scaling chosen. We therefore prefer to start with the decoupled version right away.

Remark 2. In the literature, a factor of 2 is often included in the mutation rates, selection intensities, and the time scale. The purpose of this is to make the Moran model comparable with the Wright-Fisher model, so that they share the same diffusion limit. More precisely, the factor of 2 comes from the fact that reproduction is tied to *ordered* pairs of individuals in the Moran model, but to unordered pairs in the Wright-Fisher model, cf. [6, p. 23]. Since we do not consider the Wright-Fisher model in this article, we will not stick to this convention, thus making our life easier. Let us note for completeness that a second factor of 2 may appear if populations are diploid (unlike the haploid populations considered here, they carry two copies of the genetic information per cell, which separate during reproduction, thus effectively duplicating the 'population' size).

7. Neutral genealogies. The case without selection ($s = \sigma = 0$), known as the *neutral* case in population genetics, is particularly simple. This is because reproduction rates do not depend on the type, and the reproduction and mutation processes may therefore be superimposed independently. This allows one to go beyond the 'frequency picture' considered in the previous Section, and also construct *genealogies*, i.e., answer the question: If I take a sample of m individuals from a stationary population, which is the law that governs their joint history?

We will define a *genealogy* to consist of the *genealogical tree* and the *types* along the branches; the genealogical tree, in turn, is given by the *tree topology* together with the *branch lengths* (in units of time). If a realisation of the Moran model (in forward time) is given, the genealogy can be read off immediately by starting from the sampled individuals and tracing their ancestry backward (see Fig. 5): Every time the tip of an arrow is encountered, that arrow is followed backwards, and if it joins two lines in the genealogy, these lines merge. They thus give rise to a *coalescence event*, that is, two individuals go back to a common ancestor. The crucial point now is that realisations of genealogies can be constructed without constructing realisations of the forward process first, by a two-step procedure:

Start from our m individuals and construct their genealogical tree, going backward in time and *ignoring the types*. To this end, recall that, in the graphical representation of the forward process, arrows appear at rate $1/N$ between any ordered pair of individuals (on the original time scale). Hence, arrows that lead to coalescence events appear at rate $1/N$ per ordered pair of lines in the genealogical tree. The number of lines currently in the tree is therefore a death process, which starts at m, decreases in unit steps at rate $n(n-1)/N$ if the current state is n, and ends in the absorbing state 1; at every death event, a random pair of lines is chosen to merge. The individual at the root of the tree is the *most recent common ancestor (MRCA)* of the sample. This verbal description defines the law, and an easy method of sampling, for genealogical trees. Rescaling of time turns the death rates into $n(n-1)$; the resulting stochastic process is known as the *coalescent process*, and goes back to Kingman [20, 21]. The types along the branches may then be determined in a second step, by picking the type of the most recent common ancestor from the stationary distribution (4) and running the mutation process forward in time along the genealogical tree (where, of course, both descendants inherit the parent's type at a coalescence event).

8. The ancestral selection graph. As mentioned above, the simplicity of the above construction relies on the fact that, in the neutral case, the mutation process may be superimposed independently onto the reproduction process, and hence the genealogical tree. But this independence breaks down as soon as there is selection. For this reason, the coalescent with selection has been considered intractable for more than a decade. Only within the last decade, Neuhauser and Krone [22, 23] and Donnelly and Kurtz [5] discovered constructions that overcome this problem. We will stick to the Neuhauser-Krone framework here, but the fundamental idea, in both constructions, is to again establish some kind of separation of the genealogical and the mutation processes. This is

achieved by the following trick. Start with the forward particle system, and ignore the types. Decompose the original (type-dependent) reproduction process (before rescaling of time) into two independent processes, represented by two types of arrows: 'neutral' arrows that appear at rate 1 on every line regardless of the type (as in the neutral case, hence the name); and 'selective' arrows that appear at rate s, again *on every line regardless of the type*, but with the additional convention that the latter arrows are only 'used' if they emanate from a fit individual; otherwise, they are ignored. In contrast, neutral arrows are used in either case. Put differently, neutral arrows represent *definitive birth events* for everyone, whereas selective arrows correspond to *potential birth events*, to be used by fit individuals only. If we now reintroduce the types (by defining the initial types and adding in mutation events), the resulting process is equivalent to the original Moran model with selection (forward in time). Backward in time, it allows the construction of the genealogy of a sample taken from the stationary distribution, but now three steps are required, rather than two as in the neutral case. To explain the idea, let us give a verbal description, still starting from a given realisation of the collection of arrows in a finite particle system, but without knowledge of types; later, we will free ourselves from concrete realisations, and pass to the diffusion limit.

(A1) In a first step (see Fig. 6, left panel), construct the so-called *ancestral selection graph (ASG), without types.* To this end, start from a sample of m individuals from the present population, *with the types unknown*, and *run time backward*. This is done as follows. Given the realisation of the pattern of arrows, trace back the lines; every time you meet a neutral arrow, proceed in the familiar way, thus producing a coalescence event. However, if a selective arrow is encountered, there is a potential birth event which cannot be resolved yet since the types are still unknown. Depending on whether the arrow has or has not been used, there are two possible parents: the line at the base of the arrow (the *incoming branch*), or the line at the tip (the *continuing branch*). To keep track of both possibilities, the incoming branch is attached to the graph, which results in a *branching event*. (We ignore the possibility that the incoming branch may already be contained in the graph, since the probability for this vanishes in the diffusion limit.) The resulting *branching-coalescing graph* is known as the ASG (without types); all *possible genealogical trees* are *embedded* in it. When the number of lines reaches 1 for the first time, we have found the so-called *ultimate ancestor (UA)* of the population (not to be confused with the MRCA, see below). Note, however, that, going back further in time, the line emanating from the UA will, sooner or later, branch out again; the process may be continued indefinitely.

(A2) In the second step (see Fig. 6, right panel), the type of the ultimate ancestor is drawn from the stationary distribution of the forward process. (It is not immediately obvious that the UA's type should follow this distribution; the reason it does lies in the fact that the time when the ASG reaches size 1 is independent of the type, see the arguments in [26] and [5, Thm. 8.2]. Note that, even for finite N, the stationary distribution is available in closed form [28]; but it simplifies considerably in the diffusion limit.) Starting from the UA, the mutation process is then superimposed on

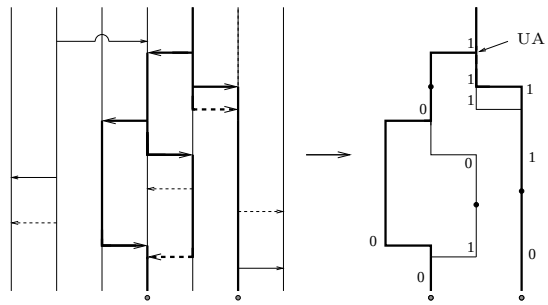


Fig. 6. Constructing the ASG. Left: Particle system with neutral arrows (solid) and selective arrows (dashed). Embedded (as bold lines) is the ASG without types, as obtained when going backward from the two 'grey' individuals sampled from the present population. Right: Starting from the UA, the mutation process is superimposed on the ASG, which may then be resolved into real (bold) and virtual (thin) branches according to Fig. 7. The collection of bold lines (along with their types) constitutes the genealogy of the grey individuals.

the graph, running *forward in time*. As usual, both descendants inherit the parent's type at a coalescence event. And now that the types are known, the branching events can be resolved as one goes along (see Fig. 7): If the incoming branch is fit, the selective birth event has indeed taken place, and the incoming branch is *parental* to the descendant individual (the one below the tip of the arrow); the continuing branch is called *virtual*. If the incoming branch is, however, unfit, then the birth event in question was fictitious, the incoming branch is virtual, and the continuing branch is the parental one. Note that, as a result, the descendant individual is unfit if and only if both the incoming and the continuing branches are unfit. The result of step (A2) is known as the *ASG with types*.

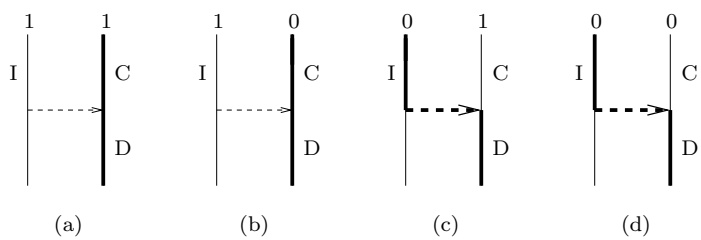


Fig. 7. Resolution of potential birth events. I incoming branch, C continuing branch, D descendant. The descendant line and its parental branch are marked bold.

(A3) In the third and final step (see again Fig. 6, right panel), the genealogy is extracted by moving backward in time once more. Starting at the sampled individuals, we trace their ancestry backwards along their parental lines identified in the previous step. The collection of all lines that are *parental to the sampled individuals* constitute the genealogical tree; they are also called *real branches*. All other lines are virtual (in particular, lines that are parental to virtual branches are themselves virtual), and are now removed from the picture to yield the genealogy. Note that the MRCA of the genealogical tree need not coincide with the UA of the ASG; it may be younger.

The diffusion limit simplifies this construction in two respects. In (A1), the probability for the incoming branch to be one that is already within the ASG vanishes in the limit. And in (A2), it provides us with the simple explicit form of the stationary distribution (4) from which to pick the UA's type. Forgetting about concrete realisations and considering that, after rescaling of time, selective arrows land on every line at rate σ, we can construct the *ASG in the diffusive limit* as the following analogue of (A1)–(A3):

(A1') Construct the number of lines in the graph as a birth-death process, starting at m, with death rate $n(n-1)$ and birth rate $n\sigma$ when there are currently n lines. From this, construct the ASG without types by choosing a random pair to coalesce at every death event, and choosing a random line to branch into two at every birth event. Stop when there is only one line left (this will almost surely happen in finite time [22, Thm. 3.2]).

(A2') Pick the type of the ultimate ancestor from the stationary distribution (4). Run the mutation process down the graph (at rates $u\nu_0$ and $u\nu_1$) and resolve branching events as you go along, according to Fig. 7.

(A3') Extract the genealogy, as in (A3).

An explicit algorithm is given in [26, Alg. 3.1].

9. Tracing back single individuals. After describing the general idea of the ASG, we now turn to the simpler situation that arises if we trace back the ancestry of *single* individuals, picked from the stationary (present) distribution, rather than the genealogies of *samples* of several individuals. For this purpose, a simplified construction is available, which was recently presented by Fearnhead [12]. For a complementary approach, which relies on diffusion theory alone (without reference to genealogies), the reader is referred to Taylor [27]; we will restrict ourselves to the genealogical approach of [12] here. From now on, we will directly work with the diffusion limit.

Even if only a single individual is sampled, its ancestral line will branch out at some stage; we will therefore still require the law for *samples*. Consider, therefore, a sample of size m (which, for the moment, comes from the present population). One may either describe it as an *ordered sample* $\boldsymbol{c} = (c_1, c_2, \ldots, c_m)$, where c_ℓ is the type of the ℓ-th individual ($1 \leq \ell \leq m$), or as an *unordered sample* $\boldsymbol{m} = (m_0, m_1)$, where m_i, $i \in S$, is the number of individuals of type i (of course, $m_0 + m_1 = m$). Since there is no spatial information in our model, the law of our sample is invariant under permutations of the individuals. More precisely, if $p^*(\boldsymbol{c})$ is the probability of getting the ordered sample $\boldsymbol{c}$ when sampling m individuals, we have $p^*(\hat{\boldsymbol{c}}) = p^*(\boldsymbol{c})$ if $\#_i(\hat{\boldsymbol{c}}) = \#_i(\boldsymbol{c})$, $i \in S$, where $\#_i(\boldsymbol{c})$ is the number of individuals of type i in $\boldsymbol{c}$; this property is known as *exchangeability* in probability theory. One could therefore work with unordered samples; for our purpose, however, it is convenient (and compatible with [12]) to work with ordered samples, but to retain a slightly abusive notation reminiscent of unordered samples. Namely, we will denote by $p(\boldsymbol{m})$ the probability of *any* ordered sample that has m_0 0-individuals and m_1 1-individuals. More precisely, $p(\boldsymbol{m})$ satisfies $p^*(\boldsymbol{c}) = p(\boldsymbol{m})$ for every $\boldsymbol{c}$ with $\#_0(\boldsymbol{c}) = m_0$ and $\#_1(\boldsymbol{c}) = m_1$. By Wright's formula (4), we have $p(\boldsymbol{m}) = \mathbb{E}(X_0^{m_0} X_1^{m_1}) = \int_{\mathcal{X}} x_0^{m_0} x_1^{m_1} f(x_0, x_1) dx_0 dx_1$, where

$\mathcal{X} := \{(x_0, x_1) \mid x_0, x_1 \geq 0 \text{ for } i \in S, x_0 + x_1 = 1\}$. For later use, let us also define

$$p(j \mid \boldsymbol{m}) := \frac{p(\boldsymbol{m} + \boldsymbol{e}_j)}{p(\boldsymbol{m})}, \quad j \in S, \tag{5}$$

where $\boldsymbol{e}_0 := (1, 0)$ and $\boldsymbol{e}_1 := (0, 1)$; in words, $p(j \mid \boldsymbol{m})$ is the probability of obtaining an individual of type j if we have already drawn a sample $\boldsymbol{m}$ and pick one additional individual from the stationary distribution.

Armed this way, let us now describe the construction of the ancestral line of a single individual. A further simplification will result from the fact that the types will be included throughout. Modifying the construction (A1')–(A3'), one proceeds in the following way (see [12] and [26] for details):

(B1) Pick an individual (i.e., a sample of size one) from the present population. Choose its type according to Wright's formula (4), i.e., $p(\boldsymbol{e}_i) = \mathbb{E}(X_i)$, $i \in S$.

(B2) Given this type, construct the ASG *including the types*. This only requires a single step because, given the initial type, the mutation process may be run backward, and, when a branching event is encountered, the type of the descendant is already known — and hence the type of the parental branch (be it 'incoming' or 'continuing'), because it coincides with the type of the descendant. This way, potential birth events can be resolved on the way back already, thus contracting steps (A1') and (A2'). In particular, this applies to the real branch (there is exactly one real branch throughout – the line parental to the single individual picked in step (B1)). What needs to be assigned at a branching event is the type of the virtual branch that appears; this can be done by using Fig. 8, and will be formalised below. Let us only conclude here that the ASG boils down to a process with states $(i; \boldsymbol{n}) \in S \times \mathbb{Z}^2_{\geq 0}$, where i denotes the type of the real branch, and $\boldsymbol{n} = (n_0, n_1)$ collects the numbers of virtual branches of type 0 and 1, respectively.

Ultimately, to determine the ancestral distribution of types analogous to that of the branching process, we will only need the real branch – but to determine its type, the virtual ones are indispensable, as will become clear in a moment.

10. Transitions and rates of the ASG, including the types. Let us now make precise step (B2) by recapitulating from [12] the definition of the ASG in terms of its transitions and the corresponding rates; for a detailed derivation, we refer the reader to [12] and [26, Alg. 3.2]. Let us only recall here that working backward in time (relative to the stationary measure p of ordered samples) is analogous to constructing the generator for the time-reversed version of a continuous-time Markov chain on a state space E, as given by $\bar{Q} = (\bar{Q}_{ij})_{i,j \in E}$, where $\bar{Q}_{ij} = \pi_i^{-1} Q_{ji} \pi_j$ if $Q = (Q_{ij})_{i,j \in E}$ is the generator of the forward process and $\pi = (\pi_i)_{i \in E}$ is the corresponding stationary distribution; cf. [24, Ch. 3.7] for some introductory material on time reversal; and compare the related time-reversed branching process in Sec. 2. Let us also note that, for ease of exposition, we will include 'empty events' in the context of mutation (as in Sec. 6).

Given that the current state is $(i; \boldsymbol{n})$ and going backward in time, five types of events can occur (cf. Fig. 8); note that $(i; \boldsymbol{n})$ corresponds to a sample $\boldsymbol{n} + \boldsymbol{e}_i$.

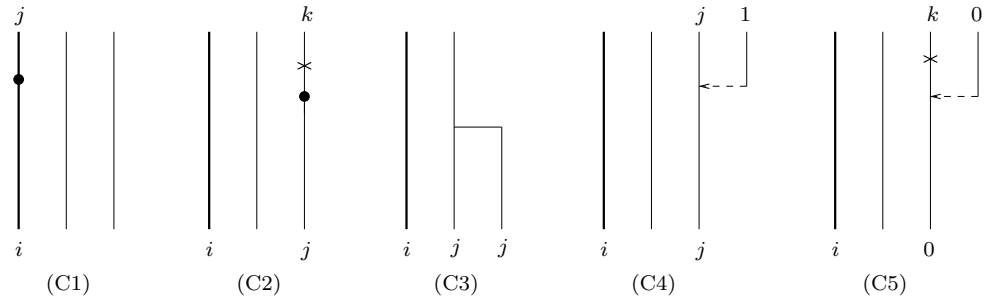


Fig. 8. Transitions of the ASG, including the types. Transitions (C1)–(C5) out of $(i;\boldsymbol{n})$ in the ASG with types. The bold line is the real branch, the thin lines represent virtual branches. Blobs mark mutation events, crosses mark branches whose types are obtained as a random draw according to (5), and may be removed. In (C4), the continuing branch is parental; in (C5), the incoming branch is parental. Backward time runs from bottom to top.

(C1) *Mutation of the real branch:* $(i;\boldsymbol{n}) \to (j;\boldsymbol{n})$, $j \in S$, occurs at rate $\theta\nu_i p(\boldsymbol{n}+\boldsymbol{e}_j)/p(\boldsymbol{n}+\boldsymbol{e}_i)$, in accordance with the above reversion recipe.

(C2) *Mutation of a virtual branch*: $(i;\boldsymbol{n}) \to (i;\boldsymbol{n}+\boldsymbol{e}_k-\boldsymbol{e}_j)$, $j,k \in S$, occurs at rate $n_j\theta\nu_j p(\boldsymbol{n}+\boldsymbol{e}_i-\boldsymbol{e}_j+\boldsymbol{e}_k)/p(\boldsymbol{n}+\boldsymbol{e}_i)$. The factor n_j comes from the fact that, in the corresponding forward transition, $(i,\boldsymbol{n})$ is the target state and contains n_j virtual branches of type j, each of which may have arisen by mutation (recall that $p(\boldsymbol{m})$ is the probability of *any ordered* sample of composition $\boldsymbol{m}$). Let us note for later use that, by (5), the rate may be rewritten as $(n_j\theta\nu_j p(\boldsymbol{n}+\boldsymbol{e}_i-\boldsymbol{e}_j)/p(\boldsymbol{n}+\boldsymbol{e}_i))\times p(k \mid \boldsymbol{n}+\boldsymbol{e}_i-\boldsymbol{e}_j)$, that is, the type of the mutated branch can be obtained as a random draw from the stationary population, given the other branches in the sample.

(C3) *Coalescence of two branches of type j, $j \in S$*: $(i;\boldsymbol{n}) \to (i;\boldsymbol{n}-\boldsymbol{e}_j)$ occurs at rate $(n_j+\delta_{ij})(n_j+\delta_{ij}-1)p(\boldsymbol{n}+\boldsymbol{e}_i-\boldsymbol{e}_j)/p(\boldsymbol{n}+\boldsymbol{e}_i)$. The combinatorial factor reflects the fact that coalescence events may involve both real and virtual branches, but can only occur between like types; for type j, there are $(n_j+\delta_{ij})(n_j+\delta_{ij}-1)$ ordered pairs of type j.

(C4) *Branching to an unfit incoming branch* (see also Fig. 7, (a) and (b)): $(i;\boldsymbol{n}) \to (i;\boldsymbol{n}+\boldsymbol{e}_1)$ occurs at rate $(n_j+\delta_{ij})\sigma p(\boldsymbol{n}+\boldsymbol{e}_i+\boldsymbol{e}_1)/p(\boldsymbol{n}+\boldsymbol{e}_i)$, for every $j \in S$. Here, the continuing branch (regardless of its type) is parental, and its type is thus the same as before the branching event.

(C5) *Branching to a fit incoming branch* (see also Fig. 7, (c) and (d)): $(i;\boldsymbol{n}) \to (i;\boldsymbol{n}+\boldsymbol{e}_k)$, $k \in S$, occurs at rate $(n_0+\delta_{i0})\sigma p(\boldsymbol{n}+\boldsymbol{e}_i+\boldsymbol{e}_k)/p(\boldsymbol{n}+\boldsymbol{e}_i) = (n_0+\delta_{i0})\sigma p(k \mid \boldsymbol{n}+\boldsymbol{e}_i)$. Here, the incoming branch is parental; the type of the continuing branch may be obtained as a random draw conditional on the remaining sample, similar to the situation in (C2).

The form of the mutation rate in (C1) is essential to understanding why we must keep track of the virtual branches at all, if we are, ultimately, only interested in the type of the real branch: the mutation process on the real branch depends on the full state of the process, including the virtual branches. Considering the real branch in isolation

would, therefore, not lead to a *Markov* process. Indeed, the achievement of the ASG construction, and further simplifications like the *common ancestor process* below, lies just in the construction of a process with a state space that is 'as small as possible', with as few transitions as possible, while still retaining the Markov property. Indeed, the lines in the ASG only represent a vanishing proportion of the population (which is infinite in the diffusion limit!); loosely speaking, they may be considered a 'representative sample'.

11. The common ancestor process. Since we are only interested in the history of one individual, genealogical trees do not matter. This allows for a further simplification. According to [12, Thm. 1], virtual branches whose types have the law of a random draw from the stationary population (conditionally on the types of the remaining sample, as in events (C2) and (C5)) may be removed; the remaining ones still perform a Markov process defined by the events and rates (C1)–(C5). The intuition behind this is the following. Let k be the type of a virtual branch in question, in a sample $(\boldsymbol{m} + \boldsymbol{e}_k)$. If k is distributed according to $p(k \mid \boldsymbol{m})$, then it contains no information about the history of the remaining sample, as suggested by the following thought experiment. Consider you draw $m + 1$ individuals from the population, but determine only the type of m of them, resulting in the (known) sample $\boldsymbol{m}$. Then $p(k \mid \boldsymbol{m})$ is the law of the type of the unknown individual. Clearly, then, the history of the known individuals is an ASG, regardless of the unknown type.

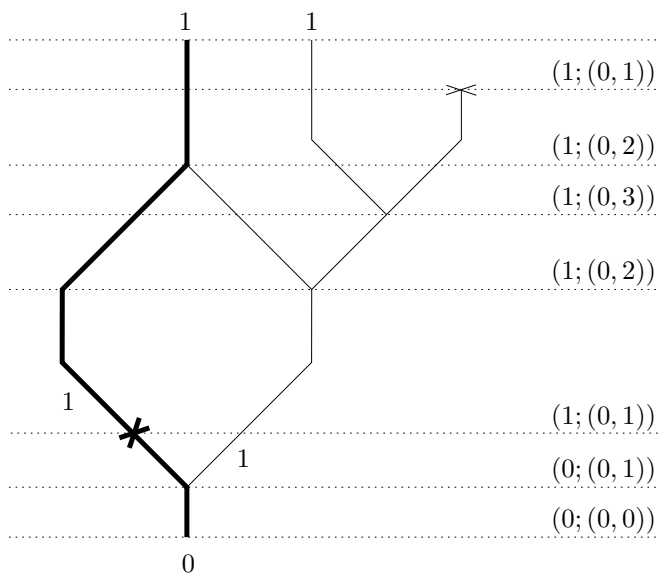


Fig. 9. The common ancestor process. The solid line is the real branch, thin lines are virtual branches, numbers on lines mark their type. Backward time runs from bottom to top. Mutation events are marked by crosses; they lead to a change of type on the real branch, but to extinction on virtual branches. Virtual branches are all unfit. The state of the process between events is given to the right.

Consider now events (C2) and (C5). Since, in both cases, the type of a virtual branch has a distribution of the form $p(k \mid \boldsymbol{m})$, it can safely be removed from the ASG without

influencing the other branches (and, in particular, the history of the ancestor). Removal of the mutated virtual branch in (C2) turns this transition into one that is indistinguishable from a coalescence event of two virtual branches, (C3). Step (C5) may be removed altogether, since removal of the newly-created virtual branch turns the transition into an 'empty' one. As a consequence, virtual branches can now only arise due to (C4) and are all unfit. Therefore, $n_0 = 0$ throughout, and we set $n := n_1$ in what follows. We thus arrive at what was termed the *common ancestor process (CAP)* in [12], and defined by its transitions out of $(i;\boldsymbol{n})$:

(P1) *Mutation of the real branch*: $(i;\boldsymbol{n}) \to (j;\boldsymbol{n})$, $j \in S$, at rate $\theta\nu_i p(\boldsymbol{n}+\boldsymbol{e}_j)/p(\boldsymbol{n}+\boldsymbol{e}_i)$;
(P2) *Removal of a virtual branch (by coalescence or mutation)*: $(i;\boldsymbol{n}) \to (i;\boldsymbol{n}-\boldsymbol{e}_1)$ at rate $[(n_1+\delta_{i1})(n_1+\delta_{i1}-1)+\theta\nu_1 n_1]p(\boldsymbol{n}+\boldsymbol{e}_i-\boldsymbol{e}_1)/p(\boldsymbol{n}+\boldsymbol{e}_i)$;
(P3) *Branching*: $(i;\boldsymbol{n}) \to (i;\boldsymbol{n}+\boldsymbol{e}_1)$ at rate $\sigma(n_1+1)p(\boldsymbol{n}+\boldsymbol{e}_i+\boldsymbol{e}_j)/p(\boldsymbol{n}+\boldsymbol{e}_i)$. (Note the typo in this rate in [12]: the factor (n_1+1) is missing.)

The process starts in state $(i;(0,0))$ with probability $p(\boldsymbol{e}_i)$. A realisation is shown in Fig. 9. Note that the above notation of the states is somewhat redundant (the support of the process is $E = S \times \{0\} \times \mathbb{Z}_{\geq 0}$), but we retain it here for compatibility with the 'full' ASG.

12. The stationary distribution of the common ancestor process. By Thm. 3 and Lemma 1 of [12], the stationary distribution a of the CAP is given by

$$a(i;\boldsymbol{n}) = A_n(i)p(\boldsymbol{n}+\boldsymbol{e}_i), \quad (i;\boldsymbol{n}) \in E, \tag{6}$$

where $A_n(0) := \prod_{j=1}^{n}\lambda_j$, $A_n(1) := A_n(0)(1-\lambda_{n+1})$, and $\lambda_j := \lim_{k\to\infty}\lambda_j^{(k)}$ $(j \geq 1)$, where the $\lambda_j^{(k)}$ $(k \geq 0)$ are defined by $\lambda_{k+1}^{(k)} = 0$, together with the recursion

$$\lambda_{j-1}^{(k)} = \frac{\sigma}{j+\theta+\sigma-(j+\theta\nu_1)\lambda_j^{(k)}}\,. \tag{7}$$

The proof is by direct, though slightly cumbersome, verification; an alternate proof, which uses advanced diffusion theory rather than the ASG, is given in [27]. But a nice interpretation is also available [12], which turns (6) into the following rule for simulating from the distribution a [12]. Choose an individual at random from the present distribution (4). If the individual is fit, call it the real branch (i.e., the ancestor). If the individual is unfit, then with probability $1-\lambda_1$ call it the ancestor; otherwise, call it a virtual branch, and pick another individual from the present distribution. Repeat this (i.e., if the j'th individual chosen is unfit, then with probability $1-\lambda_j$ it is the ancestor, otherwise draw another one) until you find the ancestor. When the ancestor has been found, the state $(i;\boldsymbol{n})$ is given by the type i of the ancestor, and the number n_1 of virtual branches (and $n_0 = 0$).

Marginalising over the virtual branches finally results in the stationary distribution of the ancestral type; more precisely, $a_i := \sum_{n\geq 0} a(i;n)$ is the stationary probability that the ancestor has type i. Clearly, this is the desired analogue of α_i in the branching process.

13. An example. Recall the two-type branching caricature of the single-peaked model (Sec. 5), and let us compare it to its Moran analogue. For the values of s and ν_0 used before, Fig. 10 shows the stationary frequencies of fit individuals in the present and ancestral population, as a function of u, and for various population sizes (that is, we take the approximating diffusion defined by the choice $\sigma = Ns$, $\theta = Nu$, for given s and u). More precisely, the Figure depicts the expected frequency of fit individuals in the present population, $p(\boldsymbol{e}_0) = \mathbb{E}(X_0)$ as given by Wright's formula; and the expected frequency of fit ancestors of individuals sampled randomly from the present population, as given by a_0. Comparing these with the corresponding stationary frequencies π_0 and α_0 in the branching process suggests that the stationary distribution of the branching process is related to another $N \to \infty$ limit of the Moran model, which differs from the diffusion limit. Indeed, the Moran model (forward in time) also has a law of large numbers: Consider a sequence of Moran models *with fixed s and u*, and increasing N, *without rescaling of time*. Denote the corresponding normalised processes by $\{X_0^{(N)}(t), X_1^{(N)}(t)\}_{t\geq 0}$, where the upper index is now introduced to denote dependence on population size. One can then show that, for $N \to \infty$, the sequence of generators of the processes $\{X_0^{(N)}(t), X_1^{(N)}(t)\}_{t\geq 0}$ converges, in a suitable sense, to the generator of the deterministic process defined by the system of ordinary differential equations

$$\dot{p}_i(t) = (R_i - \langle R, p\rangle)p_i(t) + \sum_{j:j\neq i} (p_j(t)U_{ji} - p_i(t)U_{ij}), \quad i \in S, \tag{8}$$

where $p_i(t)$ is the relative frequency of type i at time t (details will be given elsewhere).

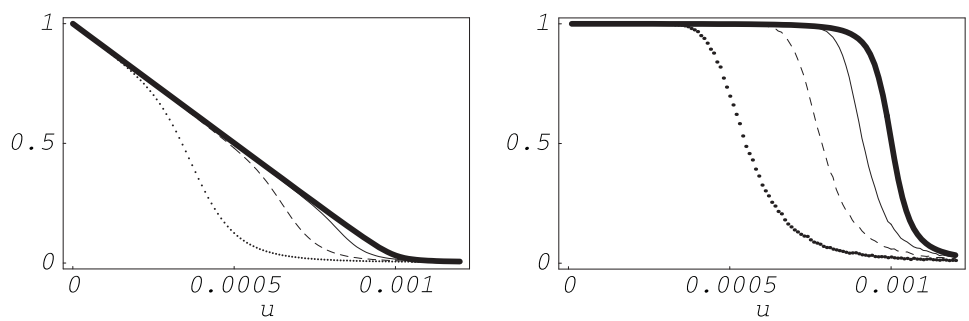


Fig. 10. The present and ancestral frequencies of the fit type in the two-type branching process and the corresponding Moran model with selection, as a function of the mutation rate. Parameters: $s = 0.001$, $\nu_1 = 0.999$. Left: Expected frequency of fit individuals in the stationary *present* distribution. Bold line: π_0, of the branching process; thin lines: $p(\boldsymbol{e}_0)$ in the Moran model, for $N = 10000$ (dotted), $N = 30000$ (dashed), and $N = 100000$ (solid line), according to Wright's formula (4). Right: Expected frequency of fit individuals in the ancestral distribution. Bold line: α_0, of the branching process; thin lines: a_0 in the Moran model, for $N = 10000$ (dotted), $N = 30000$ (dashed), and $N = 100000$ (solid line). Here, a_0 is approximated by simulating according to the recipe resulting from (6), and averaging over 10000 realisations. The λ_j required in the simulation are approximated by setting $\lambda_{500}^{(499)} = 0$, and using the recursion (7).

By Thms. 1.6.1 and 4.9.10 of [10], this implies weak convergence of the stationary distributions of $\{X_0^{(N)}(t), X_1^{(N)}(t)\}_{t\geq 0}$ to the stationary distribution of (8). But the latter, in

turn, is easily seen to coincide with the stationary distribution π of the branching process (see, e.g., [3]). (A corresponding relationship for the backward process, suggestive as it may seem, remains to be formally established.)

It is interesting to observe how the convergence of the stationary distribution manifests itself (see Fig. 10). The Moran curves follow the branching curves closely for small u and then break off; this happens at values of u that approach $u_0 = s$ with increasing population size. Therefore, the branching result is an excellent approximation even for small and moderate-sized populations, as long as mutation rates are small.

These observations complement previous studies [25, 28], which only considered the forward direction of time (they actually originated well before the ancestral selection graph). Let us note that the transition (in both branching and Moran models) becomes much steeper when the mutation rates are even more asymmetric than assumed here; the behaviour can then be interpreted as an error threshold for finite populations [28], but this is not our concern here.

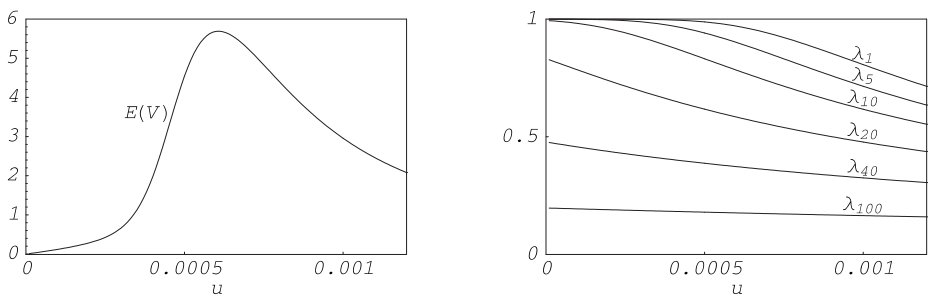


Fig. 11. The virtual branches and how they originate. The situation corresponds to the dotted curve in Fig. 10, i.e., $N = 10000$, $s = 0.001$, $\nu_1 = 0.999$. Left panel: Expected number of virtual branches, as a function of the mutation rate. The number of virtual branches, V, is simulated as in Fig. 10 (right), and averaged over 10000 realisations to obtain an approximation of its expectation, $E(V)$. Right panel: The λ_j as a function of the mutation rate, calculated as in Fig. 10 (right).

14. The virtual branches. The transition described above is intimately connected with the number of virtual branches – in fact, it is accompanied by an explosion of the expected number of virtual branches, see Fig. 11 (left panel). This can only be understood in the light of how the λ_j depend on u. As shown in Fig. 11 (right panel), λ_1 is very close to 1 at $u = 0$, and the λ_j decrease with u, as well as j. The shape of the ancestral distribution may then be understood as follows. For $N = 10000$, s, ν as in Fig. 10, and u well below 0.0005, simulation according to the 'recipe' resulting from (6) is similar to drawing from the present distribution until a fit individual is found, which is then identified as the ancestor — this is because λ_j is close to 1 for $j \leq 100$; and the expected frequency of fit individuals in the present population, $p(\boldsymbol{e}_0)$, is not too small, so that, typically, only few draws are required to find a fit individual. Therefore, the ancestors are almost all fit, although the present population is not — at the expense of an increasing number of virtuals. With increasing u (and, correspondingly, decreasing $p(\boldsymbol{e}_0)$), however, one needs

more and more trials to find a fit individual; this leads to a further increase in the number of virtuals, increasing j and decreasing λ_j. As a consequence, it becomes more and more likely to 'accept' an unfit individual as the ancestor; finally (beyond $u \sim 0.0005$), both the present and the ancestral population consist of mainly unfit individuals. But a large number of virtual branches is still present, because, for small j, the λ_j are still close to one. The number of virtual branches only decreases slowly, in accordance with the decrease of the λ_j with u.

Acknowledgements. It is our pleasure to thank Michael Baake, Paul Fearnhead, Steve Krone, and Jay Taylor for enlightening discussions and help with details, and Jay Taylor and John Wakeley for critically reading the manuscript. This work was supported by the Dutch-German Bilateral Research Group on Random Spatial Models in Physics and Biology (DFG-FOR 498).

References

[1] K. B. Athreya and P. E. Ney, *Branching Processes*, Springer, New York, 1972.

[2] E. Baake and W. Gabriel, *Biological evolution through mutation, selection, and drift: an introductory review*, in: D. Stauffer (ed.), Ann. Rev. Comp. Phys. 9, World Scientific, Singapore, 2000, 203–264; arXiv:cond-mat/9907372.

[3] E. Baake and H.-O. Georgii, *Mutation, selection and ancestry in branching models: a variational approach*, J. Math. Biol. 54 (2007), 257–303; arXiv:q-bio/0611018.

[4] P. Donnelly and T. Kurtz, *A countable representation of the Fleming-Viot measure-valued diffusion*, Ann. Prob. 24 (1996), 698–742.

[5] P. Donnelly and T. Kurtz, *Genealogical processes for Fleming-Viot models with selection and recombination*, Ann. Appl. Prob. 9 (1999), 1091–1148.

[6] R. Durrett, *Probability Models for DNA Sequence Evolution*, Springer, New York, 2002.

[7] M. Eigen, *Selforganization of matter and the evolution of biological macromolecules*, Naturwiss. 58 (1971), 465–523.

[8] M. Eigen, J. McCaskill, and P. Schuster, *The molecular quasi-species*, Adv. Chem. Phys. 75 (1989), 149–263.

[9] A. Etheridge, *Introduction to Superprocesses*, AMS, Providence, RI, 2000.

[10] S. N. Ethier and T. G. Kurtz, *Markov Processes—Characterization and Convergence*, Wiley, New York, 1986; Reprint 2005.

[11] W. Ewens. *Mathematical Population Genetics*, 2nd ed., Springer, Berlin, 2004.

[12] P. Fearnhead, *The common ancestor at a nonneutral locus*, J. Appl. Prob. 39 (2002), 38–54.

[13] H.-O. Georgii and E. Baake, *Multitype branching processes: the ancestral types of typical individuals*, Adv. Appl. Prob. 35 (2003), 1090–1110; arXiv:math/0302049.

[14] J. Hermisson, O. Redner, H. Wagner, and E. Baake, *Mutation-selection balance: Ancestry, load, and maximum principle*, Theor. Pop. Biol. 62 (2002), 9–46; arXiv:cond-mat/0202432.

[15] P. Jagers, *General branching processes as Markov fields*, Stoch. Proc. Appl. 32 (1989), 183–242.

[16] P. Jagers, *Stabilities and instabilities in population dynamics*, J. Appl. Prob. 29 (1992), 770–780.

[17] K. S. Karlin and H. M. Taylor, *A First Course in Stochastic Processes*, 2nd ed., Academic Press, San Diego, CA, 1975.
[18] J. G. Kemeny and J. L. Snell, *Finite Markov Chains*, Springer, New York, 1981.
[19] H. Kesten and B. P. Stigum, *A limit theorem for multidimensional Galton-Watson processes*, Ann. Math. Statist. 37 (1966), 1211–1233.
[20] J. F. C. Kingman, *The coalescent*, Stoch. Proc. Appl. 13 (1982), 235–248.
[21] J. F. C. Kingman, *On the genealogy of large populations*, J. Appl. Prob. 19A (1982), 27–43.
[22] S. Krone and C. Neuhauser, *Ancestral processes with selection*, Theor. Pop. Biol. 51 (1997), 210–237.
[23] C. Neuhauser and S. Krone, *The genealogy of samples in models with selection*, Genetics 145 (1997), 519–534.
[24] J. R. Norris, *Markov Chains*, Cambridge University Press, Cambridge, 1997 (reprint 1999).
[25] M. Nowak and P. Schuster, *Error thresholds of replication in finite populations: Mutation frequencies and the onset of Muller's ratchet*, J. Theor. Biol. 137 (1989), 375–395.
[26] M. Stephens and P. Donnelly, *Ancestral inference in population genetics models with selection*, Austral. New Zeal. J. Stat. 51 (2003), 901–931.
[27] J. E. Taylor, *The common ancestor process for a Wright-Fisher diffusion*, Electron. J. Probab. 12 (2007), 808–847.
[28] T. Wiehe, E. Baake, and P. Schuster, *Error propagation in reproduction of diploid organisms*, J. Theor. Biol. 177 (1995), 1–15.

STOCHASTIC MODELS IN BIOLOGICAL SCIENCES
BANACH CENTER PUBLICATIONS, VOLUME 80
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WARSZAWA 2008

REPEAT DISTRIBUTIONS FROM UNEQUAL CROSSOVERS

MICHAEL BAAKE

Fakultät für Mathematik, Universität Bielefeld
Postfach 100131, 33501 Bielefeld, Germany
E-mail: mbaake@math.uni-bielefeld.de

Abstract. It is a well-known fact that genetic sequences may contain sections with repeated units, called repeats, that differ in length over a population, with a length distribution of geometric type. A simple class of recombination models with single crossovers is analysed that result in equilibrium distributions of this type. Due to the nonlinear and infinite-dimensional nature of these models, their analysis requires some nontrivial tools from measure theory and functional analysis, which makes them interesting also from a mathematical point of view. In particular, they can be viewed as quadratic, hence nonlinear, analogues of Markov chains.

1. Introduction. Recombination is a by-product of (sexual) reproduction, which leads to the mixing of parental genes by exchanging genes (or sequence parts) between homologous chromosomes (or DNA strands). This is achieved through an alignment of the corresponding sequences, along with crossover events that lead to a reciprocal exchange of the induced segments. In this process, an imperfect alignment may result in sequences that differ in length from the parental ones; this is known as *unequal crossover* (UC). Imperfect alignment is facilitated by the presence of repeated elements (as is observed within some rDNA sequences, compare [10]), and is believed to be an important driving mechanism for the evolution of the corresponding copy number distribution. The perhaps best studied case of repeated elements concerns microsatellites, see [9] and references given there for a summary. An important observation is that, within a population, the copy numbers vary, and often (at least approximately) follow a distribution of geometric type (meaning a geometric distribution or a finite convolution product thereof), see [9, 13, 3] and references therein for some experimental examples and findings.

2000 *Mathematics Subject Classification*: Primary 92D25, 92D10; Secondary 60B10, 46B50.

Key words and phrases: recombination, crossover dynamics, quadratic operators, stationary distributions, generating functions.

It is a pleasure to thank E. Baake and O. Redner for their cooperation, and C. Bank for carefully reading the manuscript.

The paper is in final form and no version of it will be published elsewhere.

 © Instytut Matematyczny PAN, 2008

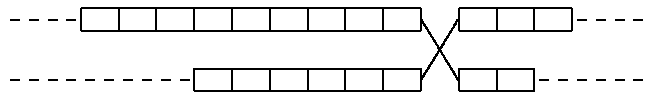


Fig. 1. Snapshot after an unequal crossover event as described in the text. Rectangles denote the relevant blocks, while the dashed lines indicate possible extensions with other elements that are disregarded here.

The microsatellites themselves may follow an evolutionary course independent of each other and thus give rise to evolutionary innovation. For a detailed discussion of these topics, see [9, 23] and references therein; for a brief introduction to molecular evolution, see also [8], or [7, 25] for a thorough overview. In this paper, which is mainly based on previous work by Redner [19, 18], we shall focus on the distribution of the copy numbers only, and disregard further aspects of the possible evolution of the repeated units themselves. We rather aim at analysing some simple models in order to understand the observed copy number or repeat distributions. Moreover, we are primarily interested in models that preserve the mean copy number, though our setting will be adequate to accommodate also more general models. In view of possible applications to systems where the copy number (slowly) changes with time, it seems natural to set up a frame that can cope with such a situation as well.

In the entire model class to be described below, one considers individuals whose genetic sequences contain a section with repeated units. These may vary in number, $i \in \mathbb{N}_0 = \{0, 1, 2, 3, \ldots\}$, where $i = 0$ is explicitly allowed and corresponds to no unit being present (yet). The composition of these sections (with respect to mutations that might have occurred) and the rest of the sequence are ignored here, as are details of the actual alignment process (e.g., whether partial loops of longer pieces are formed in order not to disturb the alignment outside the repeat region), see also [4] for a first discussion of possible models in this direction.

In the course of time, recombination events take place in which a random pair of individuals is formed and their respective sections are randomly aligned, possibly imperfectly with 'overhangs'. Then, both sequences are cut at a common position between two building blocks and their right (or left) fragments are interchanged. This so-called *unequal crossover* is schematically depicted in Figure 1. Obviously, the total number of relevant units is conserved in each event. While this is clearly a stochastic process, it is nevertheless interesting to investigate its deterministic limit, at least as a first step towards a better general understanding of this model class. To contribute to this first step, and to summarise what has been done in this direction so far, is the main aim of this contribution.

2. Description of the deterministic limit. As a first step for the analysis of crossover dynamics, we assume the population size to be (effectively) infinite, i.e., large enough so that random fluctuations may be neglected (finite populations will briefly be mentioned later on). We write $\mathcal{M}(X)$ for the (finite) measures on a space X, denote the restriction to positive measures by a superscript $+$, and indicate a restriction to measures of total

variation r by a corresponding subscript (see [20] or [27] for a short summary of the measure theory needed here). Then, the distribution of the copy numbers over our population is described by a probability measure (or vector) $\boldsymbol{p} \in \mathcal{M}_1^+(\mathbb{N}_0)$, which we identify with an element $\boldsymbol{p} = (p_k)_{k\in\mathbb{N}_0}$ in the appropriate subset of $\ell^1(\mathbb{N}_0)$. Since we do not consider any genotype space other than $\mathbb{N}_0$ in this article, reference to it will be omitted in the sequel, so we write ℓ^1 instead of $\ell^1(\mathbb{N}_0)$ from now on. These spaces are complete in the metric derived from the usual ℓ^1 norm, which is the same as the total variation norm here. The metric is denoted by

$$d(\boldsymbol{p},\boldsymbol{q}) = \|\boldsymbol{p}-\boldsymbol{q}\|_1 = \sum_{k\geq 0} |p_k - q_k|. \tag{1}$$

Let us consider the above process (as well as various more general ones) on the level of the induced dynamics on the probability measures (i.e., in the infinite population limit mentioned above). With the notation just introduced, the dynamics can be described by means of the *recombinator*

$$\mathcal{R}(\boldsymbol{p})_i := \frac{1}{\|\boldsymbol{p}\|_1} \sum_{j,k,\ell\geq 0} T_{ij,k\ell}\, p_k\, p_\ell. \tag{2}$$

Here, $T_{ij,k\ell} \geq 0$ denotes the probability that a pair (k,ℓ) turns into (i,j), so, for normalisation, we require

$$\sum_{i,j\geq 0} T_{ij,k\ell} = 1, \qquad \text{for all } k,\ell \in \mathbb{N}_0. \tag{3}$$

The factor $p_k\, p_\ell$ in (2) describes the probability that a pair (k,ℓ) is formed, i.e., we assume that two individuals are chosen independently from the population. We assume further that, for all i,j,k,ℓ,

$$T_{ij,k\ell} = T_{ji,k\ell} = T_{ij,\ell k}, \tag{4}$$

i.e., that $T_{ij,k\ell}$ is symmetric with respect to both index pairs, which is reasonable and follows from the corresponding symmetry of the underlying process, compare Figure 1. Then, the summation over j in (2) represents the breaking-up of the pairs after the recombination event. These two ingredients (symmetry and summation) lead to the quadratic nature of the iteration process, see below for more and [14, 15] for the appearance of similar types of equations in a different class of biological models.

Condition (3) and the presence of the prefactor $1/\|\boldsymbol{p}\|_1$ in the defining Eq. (2) make $\mathcal{R}$ norm non-increasing, i.e., $\|\mathcal{R}(\boldsymbol{x})\|_1 \leq \|\boldsymbol{x}\|_1$, and positive homogeneous of degree 1, i.e., $\mathcal{R}(a\boldsymbol{x}) = |a|\mathcal{R}(\boldsymbol{x})$, for $\boldsymbol{x} \in \ell^1$ and $a \in \mathbb{R}$. Furthermore, $\mathcal{R}$ is a positive operator with $\|\mathcal{R}(\boldsymbol{x})\|_1 = \|\boldsymbol{x}\|_1$ for all positive elements $\boldsymbol{x} \in \ell^1$. Thus, it is guaranteed that $\mathcal{R}$ maps $\mathcal{M}_r^+$, the space of positive measures of total variation r, into itself. This subspace is complete in the topology induced by the norm $\|.\|_1$, i.e., by the metric d from (1). (For $r = 1$, the prefactor on the right hand side of (2) is redundant, but improves numerical stability of an iteration with the nonlinear mapping $\mathcal{R}$.)

Given an initial configuration $\boldsymbol{p}_0 = \boldsymbol{p}(0)$, the dynamics may be taken in discrete time steps, with subsequent generations,

$$\boldsymbol{p}(t+1) = \mathcal{R}(\boldsymbol{p}(t)), \qquad t \in \mathbb{N}_0. \tag{5}$$

This iteration reflects the following: due to random mating, it is sufficient to consider the dynamics at the level of the single strands, which will be combined into pairs again randomly in the next generation, according to the Hardy-Weinberg equilibrium [7].

Our treatment of this case will be set up in a way that also allows for a generalisation of the results to the analogous process in continuous time, where generations are overlapping,

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\,\boldsymbol{p}(t) = \varrho\,(\mathcal{R}-\mathbb{1})(\boldsymbol{p}(t)), \qquad t \geq 0. \tag{6}$$

This reflects what is called *instant mixing*, i.e., the instantaneous formation of pairs, their recombination and separation. In other words, the actual duration of the diplophase (or "paired phase") is neglected, which is an approximation that is justified as long as recombination is rare on the time scale of an individual life span.

Obviously, the (positive) parameter ϱ in (6) only leads to a rescaling of the time t. We therefore choose $\varrho = 1$ without loss of generality. Furthermore, it is easily verified that the fixed points of (5) are in one-to-one correspondence with the equilibria of (6). (In the sequel, we use the term fixed point for both discrete and continuous dynamics.)

As mentioned above, our main interest at present is in processes that conserve the total copy number in each event, i.e., $T_{ij,k\ell} > 0$ for $i+j = k+\ell$ only. More general scenarios are possible, and also interesting, but already the concept of an equilibrium gets rather involved, whence we do not go into further details here. Together with the normalisation (3) and the symmetry condition from above, this yields

$$\sum_{i,j\geq 0} i\,T_{ij,k\ell} = \sum_{i,j\geq 0} \frac{i+j}{2}\,T_{ij,k\ell} = \sum_{i,j\geq 0} T_{ij,k\ell}\,\frac{k+\ell}{2} = \frac{k+\ell}{2}, \tag{7}$$

the second equality of which is an alternative condition that can replace the strict preservation of the copy number as follows.

Lemma 1. *Let $\mathcal{R}$ be defined by* (2)*, with $T_{ij,k\ell} \geq 0$ subject to the normalisation* (3) *and the symmetry conditions* (4)*. If also the second equality in* (7) *is satisfied, for all $k,\ell \in \mathbb{N}_0$, the mean copy number in the population is preserved.*

Proof. This is a simple calculation,

$$\sum_{i\geq 0} i\,\mathcal{R}(\boldsymbol{p})_i = \sum_{i,j,k,\ell\geq 0} i\,T_{ij,k\ell}\,p_k\,p_\ell = \sum_{k,\ell\geq 0} \frac{k+\ell}{2}\,p_k\,p_\ell = \sum_{k\geq 0} k\,p_k,$$

which shows the claim, provided that the mean $\mathfrak{m} := \sum_i i p_i$ is well-defined. ∎

From now on, we use the symbol $\mathfrak{m}$ for the mean, in order not to confuse it with summation indices and the like.

3. Markov chains for comparison. Let us take a brief detour to look at the linear counterpart, a countable state Markov chain, in the deterministic limit of the forward equation for the time evolution of its probability distribution. To this end, consider again probability vectors $\boldsymbol{p}$ on $\mathbb{N}_0$ and define

$$\mathfrak{M}(\boldsymbol{p})_k := \sum_{\ell=0}^{\infty} M_{k\ell}\,p_\ell,$$

for $k \in \mathbb{N}_0$, where all $M_{k\ell} \geq 0$ together with $\sum_{k=0}^{\infty} M_{k\ell} = 1$ for all $\ell \in \mathbb{N}_0$. This also makes the above sums well-defined on all elements of ℓ^1. Note that the matrix $M = (M_{k\ell})_{k,\ell\in\mathbb{N}_0}$ is transposed in comparison with the standard convention for Markov chains [21], because we are using it here in a dynamical systems context, with action of the matrix to the column vector on the right. The time evolution now either reads

$$\boldsymbol{p}(t+1) = \mathfrak{M}(\boldsymbol{p}(t)) \qquad \text{(in discrete time)} \tag{8}$$

or

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{p} = (\mathfrak{M} - \mathbb{1})(\boldsymbol{p}) \qquad \text{(in continuous time)}, \tag{9}$$

where the rate constant is again assumed to be 1, compare the remark after Eq. (6).

The iteration of the discrete version (8) on ℓ^1 is well-defined, while uniqueness of the solution of the initial value problem for the continuous time counterpart (9) on the same space follows from its global Lipschitz property,

$$\|\mathfrak{M}(\boldsymbol{p}) - \mathfrak{M}(\boldsymbol{q})\|_1 = \|\mathfrak{M}(\boldsymbol{p}-\boldsymbol{q})\|_1 \leq \sum_{k,\ell\geq 0} M_{k\ell}\,|p_\ell - q_\ell| = \sum_{\ell\geq 0} |p_\ell - q_\ell| = \|\boldsymbol{p}-\boldsymbol{q}\|_1,$$

which holds for all $\boldsymbol{p}, \boldsymbol{q} \in \ell^1$. The properties of the matrix M guarantee that the positive cone as well as the simplex of probability vectors are preserved in forward time. Consequently, one can consider (8) and (9) as dynamical systems on ℓ^1. As the latter is a Banach space of infinite dimension, the unit ball is no longer compact in the norm topology, whence some extra care is needed for the results.

As before, fixed points of (8) line up with equilibria of (9), so that we speak of fixed points in both cases. Their existence is provided by Perron-Frobenius theory for countable state Markov matrices, see [12, Ch. 7.1] or [21, Ch. 5] for a detailed account. Irreducibility, aperiodicity and primitivity are defined as in the finite-dimensional case without difficulty. However, for meaningful results on eigenvalues and eigenvectors, one additionally needs the concept of *recurrence*, see [12, p. 197 f.] for a nice summary.

The Perron value λ emerges from the radius of convergence, ρ, of the power series $T(z) = \sum_{n\geq 0}(zM)^n$ via $\rho = 1/\lambda$. Clearly, we have $\rho \geq 1$ for a Markov matrix. If one diagonal entry (and then any) of $T(z)$ diverges at 1 (so that $\rho = 1$ in this case, compare [21, Thm. 6.6]), the countable state Markov matrix M is called *recurrent*, where the behaviour of the diagonal element $T(z)_{ii}$, as $z \to 1$, reflects the expected number of recurrences to i, which is infinite in this case. Moreover, a unique normalised and strictly positive (right) eigenvector $\boldsymbol{p} \in \mathcal{M}_1^+$ exists with $\mathfrak{M}(\boldsymbol{p}) = \boldsymbol{p}$, see [21, Thm. 5.4]. This probability vector has the meaning of the unique equilibrium distribution and is the desired fixed point of the dynamics.

Assume for a moment, in addition to the above conditions on $\mathfrak{M}$, that

$$\sum_{i\geq 0} iM_{ij} = j, \quad \text{for all } j. \tag{10}$$

As before, this is a sufficient condition for the mean to be preserved under the dynamics, because one has

$$\sum_{i\geq 0} i\,\mathfrak{M}(\boldsymbol{p})_i = \sum_{i\geq 0}\sum_{j\geq 0} iM_{ij}p_j = \sum_{j\geq 0}\sum_{i\geq 0} iM_{ij}p_j = \sum_{j\geq 0} jp_j,$$

with the interchange of summation being permissible due to absolute convergence of the sums involved, provided that $\mathfrak{m} = \sum_j j p_j$ exists. However, a condition of type (10) is usually too restrictive for a linear system, wherefore we do not impose it here. As we shall see, the mean copy number can be preserved without it.

A probability vector $\boldsymbol{p}$ is called *reversible* for $\mathfrak{M}$ when the detailed balance equation

$$M_{k\ell}\, p_\ell = M_{\ell k}\, p_k \tag{11}$$

holds for all $k, \ell \in \mathbb{N}_0$. An important consequence is that any reversible $\boldsymbol{p}$ is automatically a fixed point of $\mathfrak{M}$:

$$\mathfrak{M}(\boldsymbol{p})_k = \sum_{\ell \geq 0} M_{k\ell}\, p_\ell = \sum_{\ell \geq 0} M_{\ell k}\, p_k = p_k .$$

Reversibility often provides a simpler way to actually calculate a specific fixed point than the defining matrix eigenvalue equation.

Since the Perron-Frobenius eigenvalue λ need not be isolated in the spectrum of M, the convergence properties are more subtle than in the finite-dimensional situation. Under certain extra conditions (e.g., if λ *is* isolated), the time evolution of an arbitrary initial condition converges exponentially fast towards the fixed point. However, when the matrix M is not only recurrent, but also positive recurrent, one has at least convergence of the discrete iteration, see [21] for details. Here, positive recurrence means that the expected time for a return to the state i is *finite*, which is clearly stronger than mere recurrence.

The standard geometric distribution with parameter $\alpha \in (0,1)$ is a discrete probability distribution on $\mathbb{N}_0$, defined by the probability vector $\boldsymbol{p}$ with

$$p_n := \alpha(1-\alpha)^n, \quad \text{for } n \in \mathbb{N}_0. \tag{12}$$

Clearly, $p_n > 0$ and $\sum_{n\geq 0} p_n = 1$, while $\mathfrak{m} = \sum_{n\geq 0} n p_n = (1-\alpha)/\alpha$, so that $\alpha = 1/(\mathfrak{m}+1)$. If we define the matrix $M = (M_{ij})_{i,j\geq 0}$ by $M_{ij} = p_i$, one has

$$(M\boldsymbol{p})_i = \sum_j M_{ij} p_j = p_i \sum_j p_j = p_i ,$$

so that $M\boldsymbol{p} = \boldsymbol{p}$. One clearly has $M^n = M$ for all $n \in \mathbb{N}$. Consequently, each entry of $T(z)$ is a geometric series of the form $M_{ij}(1 + z + z^2 + \ldots)$, which thus diverges at $z = 1$. In particular, M is (positive) recurrent.

The matrix M does not satisfy Eq. (10). Nevertheless, the mean copy number is preserved in the following sense. Let $\boldsymbol{a}$ be an arbitrary probability vector with mean $\mathfrak{m}$, and $\boldsymbol{p}$ the geometric distribution according to (12) with the same mean. With the corresponding matrix M, one then finds

$$\sum_{i,j} i M_{ij} a_j = \sum_{i,j} i p_i a_j = \sum_i i p_i \sum_j a_j = \mathfrak{m},$$

which results in the mean preservation, provided one starts with an initial condition $\boldsymbol{a}$ of mean $\mathfrak{m}$. Otherwise, the iteration maps $\boldsymbol{a}$ to an image of mean $\mathfrak{m}$ in the first step, and preserves $\mathfrak{m}$ in all subsequent iterations.

Further eigenvectors of M are given by $\boldsymbol{q}^{(\ell)} := \boldsymbol{e}_0 - \boldsymbol{e}_\ell$ for $\ell \in \mathbb{N}$, where $\boldsymbol{e}_i$ is the standard basis vector with 1 in coordinate i and 0 otherwise. All these extra vectors belong to the eigenvalue 0, which is the only other eigenvalue of M. In fact, M is diagonalisable,

and it is not difficult to see that an arbitrary vector $\boldsymbol{a} = (a_0, a_1, a_2, \ldots) \in \ell^1$ can be written as a convergent expansion, $\boldsymbol{a} = \beta\boldsymbol{p} + \sum_{\ell\geq 1}(\beta p_\ell - a_\ell)\boldsymbol{q}^{(\ell)}$, where $\beta = \sum_{i\geq 0} a_i$. Consequently, the chosen eigenvectors of M form a basis of ℓ^1. If $U = (\boldsymbol{p}, \boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \ldots)$ denotes the matrix that columnwise consists of the eigenvectors of M, one has

$$M = U\,\mathrm{diag}(1, 0, 0, \ldots)\,U^{-1},$$

which makes the relation $M^n = M$ for $n \in \mathbb{N}$ particularly transparent. Moreover, one sees that M commutes with all matrices N of the form $N = UAU^{-1}$ where A has the block form

$$A = \begin{pmatrix} a & \mathbf{0}^t \\ \mathbf{0} & A' \end{pmatrix}$$

with an arbitrary matrix A'. Restricting N so that $M + N$ is still Markov, one can find multi-parameter families of Markov matrices that share the given stationary geometric distribution $\boldsymbol{p}$. The same stationary probability vector $\boldsymbol{p}$ can thus arise from many other Markov chains as well.

Let us now return to the bilinear counterpart to see which of these structural properties possess an analogue, and to describe the setting of our later analysis.

4. General structure of the bilinear system. Consider the crossover dynamics as defined by (2). Let us begin by stating the following general fact.

Proposition 1. *If the recombinator $\mathcal{R}$ of* (2) *satisfies the normalisation conditions* (3)*, one has the global Lipschitz condition*

$$\|\mathcal{R}(\boldsymbol{x}) - \mathcal{R}(\boldsymbol{y})\|_1 \leq C\|\boldsymbol{x} - \boldsymbol{y}\|_1,$$

with constant $C = 3$ on ℓ^1, respectively $C = 2$ if $\boldsymbol{x}$, $\boldsymbol{y} \in \mathcal{M}_r$.

Proof. Let $\boldsymbol{x}$, $\boldsymbol{y} \in \ell^1$ be non-zero (otherwise the statement is trivial). Then, one has

$$\begin{aligned}
\|\mathcal{R}(\boldsymbol{x}) - \mathcal{R}(\boldsymbol{y})\|_1 &= \sum_{i\geq 0}\Bigl|\sum_{j,k,\ell\geq 0} T_{ij,k\ell}\Bigl(\frac{x_k\,x_\ell}{\|\boldsymbol{x}\|_1} - \frac{y_k\,y_\ell}{\|\boldsymbol{y}\|_1}\Bigr)\Bigr| \\
&\leq \sum_{k,\ell\geq 0}\Bigl|\frac{x_k\,x_\ell}{\|\boldsymbol{x}\|_1} - \frac{y_k\,y_\ell}{\|\boldsymbol{y}\|_1}\Bigr| \sum_{i,j\geq 0} T_{ij,k\ell} = \sum_{k,\ell\geq 0}\Bigl|\frac{x_k\,x_\ell}{\|\boldsymbol{x}\|_1} - \frac{x_k\,y_\ell}{\|\boldsymbol{x}\|_1} + \frac{x_k\,y_\ell}{\|\boldsymbol{x}\|_1} - \frac{y_k\,y_\ell}{\|\boldsymbol{y}\|_1}\Bigr| \\
&\leq \sum_{k,\ell\geq 0}\Bigl(\frac{|x_k|}{\|\boldsymbol{x}\|_1}|x_\ell - y_\ell| + |y_\ell|\Bigl|\frac{x_k}{\|\boldsymbol{x}\|_1} - \frac{y_k}{\|\boldsymbol{y}\|_1}\Bigr|\Bigr) = \|\boldsymbol{x} - \boldsymbol{y}\|_1 + \frac{\bigl\|\|\boldsymbol{y}\|_1\boldsymbol{x} - \|\boldsymbol{x}\|_1\boldsymbol{y}\bigr\|_1}{\|\boldsymbol{x}\|_1}.
\end{aligned}$$

The last term becomes

$$\frac{1}{\|\boldsymbol{x}\|_1}\bigl\|\|\boldsymbol{y}\|_1\boldsymbol{x} - \|\boldsymbol{x}\|_1\boldsymbol{y}\bigr\|_1 = \frac{1}{\|\boldsymbol{x}\|_1}\bigl\|(\|\boldsymbol{y}\|_1 - \|\boldsymbol{x}\|_1)\boldsymbol{x} + \|\boldsymbol{x}\|_1(\boldsymbol{x} - \boldsymbol{y})\bigr\|_1 \leq 2\|\boldsymbol{x} - \boldsymbol{y}\|_1,$$

from which $\|\mathcal{R}(\boldsymbol{x}) - \mathcal{R}(\boldsymbol{y})\|_1 \leq 3\|\boldsymbol{x} - \boldsymbol{y}\|_1$ follows for $\boldsymbol{x}$, $\boldsymbol{y} \in \ell^1$. If $\boldsymbol{x}$, $\boldsymbol{y} \in \mathcal{M}_r$, one has $\|\boldsymbol{x}\|_1 = \|\boldsymbol{y}\|_1$ and the above calculation simplifies to $\|\mathcal{R}(\boldsymbol{x}) - \mathcal{R}(\boldsymbol{y})\|_1 \leq 2\|\boldsymbol{x} - \boldsymbol{y}\|_1$. ∎

In continuous time, this is a sufficient condition for the existence of a unique solution of the initial value problem (6), compare [1, Thms. 7.6 and 10.3].

It is instructive to generalise the notion of reversibility. We call a probability vector $\boldsymbol{p} \in \mathcal{M}_1^+$ *reversible* for a recombinator $\mathcal{R}$ of the form (2) if, for all $i, j, k, \ell \geq 0$,

$$T_{ij,k\ell}\, p_k\, p_\ell = T_{k\ell,ij}\, p_i\, p_j. \tag{13}$$

Though this set of equations for detailed balance is much more restrictive than its linear counterpart in Eq. (11), the relevance of this concept is evident from the following property.

LEMMA 2. *If $\boldsymbol{p} \in \mathcal{M}_1^+$ is reversible for $\mathcal{R}$, it is also a fixed point of $\mathcal{R}$.*

Proof. Assume $\boldsymbol{p}$ to be reversible for $\mathcal{R}$. Then, by (3),

$$\mathcal{R}(\boldsymbol{p})_i = \sum_{j,k,\ell \geq 0} T_{ij,k\ell}\, p_k\, p_\ell = \sum_{j,k,\ell \geq 0} T_{k\ell,ij}\, p_i\, p_j = p_i \sum_{j \geq 0} p_j = p_i,$$

for all $i \in \mathbb{N}_0$, which shows the claim. ∎

Returning to the original question of the existence of fixed points, we now recall the following facts, compare [6, 22] for details and proofs, which are needed for some general statements in the fixed point discussion.

PROPOSITION 2 ([27, Cor. to Thm. V.1.5]). *Assume the sequence $\bigl(\boldsymbol{p}^{(n)}\bigr)$ in $\mathcal{M}_1^+$ to converge in the weak-$*$ topology (i.e., pointwise, or vaguely) to some $\boldsymbol{p} \in \mathcal{M}_1^+$, i.e.,*

$$\lim_{n\to\infty} p_k^{(n)} = p_k \quad \text{for all } k \in \mathbb{N}_0, \qquad \text{with } p_k \geq 0 \text{ and } \textstyle\sum_{k\geq 0} p_k = 1.$$

Then, it also converges weakly (in the probabilistic sense) and in total variation, i.e., $\lim_{n\to\infty} \|\boldsymbol{p}^{(n)} - \boldsymbol{p}\|_1 = 0$. ∎

Recall from [6] that a set of measures $\mathcal{M} \subset \mathcal{M}_1^+$ is called *tight* when, for every $\varepsilon > 0$, there is an $m \in \mathbb{N}_0$ such that $\sum_{k\geq m} p_k < \varepsilon$, simultaneously for all $\boldsymbol{p} \in \mathcal{M}$. This is a uniformity condition which serves as a condition for the compactness needed later on.

PROPOSITION 3. *Assume that the recombinator $\mathcal{R}$ from (2) satisfies the normalisation (3) and possesses a convex, weak-$*$ closed invariant set $\mathcal{M} \subset \mathcal{M}_1^+$, i.e., $\mathcal{R}(\mathcal{M}) \subset \mathcal{M}$, that is tight. Then, $\mathcal{R}$ has a fixed point in $\mathcal{M}$.*

Proof. Prohorov's theorem [22, Thm. III.2.1] states that tightness and relative compactness in the weak-$*$ topology are equivalent (see also [6, Chs. 1.1 and 1.5]). In our case, $\mathcal{M}$ is tight and weak-$*$ closed, therefore, due to Proposition 2, norm compact. Further, $\mathcal{M}$ is convex by assumption, and $\mathcal{R}$ is (norm) continuous by Proposition 1. Thus, the claim follows from the Leray–Schauder–Tychonov fixed point theorem [20, Thm. V.19]. ∎

For several explicit models, we shall see that such compact invariant subsets indeed exist. On the other hand, once again due to the infinite-dimensional nature of the dynamical system, their identification and use for the various proofs is essential.

5. Takahata's model. An early and now classic example was given by Takahata [24]. In our terminology, he used a recombinator based upon the transition probabilities

$$T_{ij,k\ell} := \frac{1}{k+\ell+1}\, \delta_{i+j,k+\ell}, \tag{14}$$

for $i,j,k,\ell \in \mathbb{N}_0$. Observing $\mathrm{card}\{(i,j) \mid i,j \in \mathbb{N}_0,\, i+j = k+\ell\} = k+\ell+1$, it is clear that T just describes a recombination with uniform distribution of the copy number pairs (k,ℓ) on each (finite) block of possibilities with $k+\ell$ fixed. One can also check, via Eq. (7) and Lemma 1, that the mean $\mathfrak{m}$ is preserved. On the basis of Eq. (14), the action of the recombinator from Eq. (2) on probability vectors $\boldsymbol{p}$ simplifies to

$$\mathcal{R}(\boldsymbol{p})_i = \sum_{\substack{k,\ell \geq 0 \\ k+\ell \geq i}} \frac{1}{k+\ell+1}\, p_k p_\ell . \tag{15}$$

Though this model is mathematically rather transparent, it lacks a good intuitive justification on the level of the biological processes. Nevertheless, its properties seem to be in acceptable agreement with at least some of the observations, compare [13, 3], though other results, as those shown in [9], indicate that also other types of equilibria appear in experiment.

Proposition 4. *The probability vector $\boldsymbol{p}$ defined by*

$$p_n = \frac{1}{\mathfrak{m}+1}\left(\frac{\mathfrak{m}}{\mathfrak{m}+1}\right)^n, \quad n \in \mathbb{N}_0,$$

is a reversible equilibrium with mean $\mathfrak{m}$ for the dynamics based on T of (14).

Proof. Using standard identities with geometric series and their derivatives, it is easy to check that $\boldsymbol{p}$ indeed defines a probability vector on $\mathbb{N}_0$ with mean $\mathfrak{m}$. Detailed balance follows from a simple calculation,

$$\begin{aligned} T_{ij,k\ell}\, p_k p_\ell &= \frac{\delta_{i+j,k+\ell}}{k+\ell+1}\left(\frac{1}{\mathfrak{m}+1}\right)^2\left(\frac{\mathfrak{m}}{\mathfrak{m}+1}\right)^{k+\ell} \\ &= \frac{\delta_{k+\ell,i+j}}{i+j+1}\left(\frac{1}{\mathfrak{m}+1}\right)^2\left(\frac{\mathfrak{m}}{\mathfrak{m}+1}\right)^{i+j} = T_{k\ell,ij}\, p_i p_j , \end{aligned}$$

thus completing the claim by means of Lemma 2. ∎

These equilibria are geometric distributions as also discussed above in the Markov context. However, in view of some experimental findings reported in [9] and further arguments put forward in [23], one would like to see an initial rise, and perhaps also a maximum in the vicinity of $n \approx \mathfrak{m}$. One should note that measurements often skip the entries for small copy numbers (which seem to be rather unreliable), so that a graph with a power law decay need not indicate the absence of some (weak) form of a maximum. As the methods for the further analysis of Takahata's model are similar to what we need later on for alternative models, we first continue to investigate Takahata's model.

Theorem 1. *If the initial condition, with mean $\mathfrak{m}$, satisfies a certain tightness condition* ($\limsup_{k\to\infty} \sqrt[k]{p_k(0)} < 1$), *the dynamics, both in discrete and in continuous time, converges to the equilibrium vector $\boldsymbol{p}$ from Proposition* 4, *with* $\lim_{t\to\infty} \|\boldsymbol{p}(t) - \boldsymbol{p}\|_1 = 0$.

The proof of this theorem, quite appropriately for the present context, uses an approach via generating functions and then relies on Banach's contraction principle. It requires several preparatory steps.

Let α and δ be fixed, with $0 < \alpha \leq \delta < \infty$, and consider the space

$$X_{\alpha,\delta} := \{\boldsymbol{a} = (a_k)_{k\in\mathbb{N}_0} \mid a_0 = 1,\ a_1 = \alpha,\ \text{and } 0 \leq a_k \leq \delta^k \text{ for all } k \geq 2\}. \tag{16}$$

If equipped with the metric

$$d(\boldsymbol{a}, \boldsymbol{b}) = \sum_{k\geq 0} d_k \, |a_k - b_k|, \tag{17}$$

where $d_k = (\gamma/\delta)^k$ for some $0 < \gamma < \frac{1}{3}$, the space $X_{\alpha.\delta}$ is compact [19, Prop. 5].

Let us define a new vector, $b(\boldsymbol{p})$, for suitable $\boldsymbol{p}$, by

$$b(\boldsymbol{p})_k := \sum_{\ell\geq k} \binom{\ell}{k} p_\ell, \tag{18}$$

which is certainly well-defined for all $\boldsymbol{p}$ with $\limsup_{k\to\infty} \sqrt[k]{p_k} < 1$, by an application of [19, Prop. 6]. This proposition also clarifies the connection with the space $X_{\alpha,\delta}$ for suitable parameters α and δ. As we shall see, $X_{\alpha,\delta}$ is an example of a compact, convex space that is invariant under the recombinator dynamics. It is easy to check that one has $b(\boldsymbol{p})_0 = 1$ and $b(\boldsymbol{p})_1 = \mathfrak{m}$, so that we need $X_{\alpha,\delta}$ with $\alpha = \mathfrak{m}$ and $\delta \geq \mathfrak{m}$.

Lemma 3. *For any $\boldsymbol{p}$ with $\limsup_{k\to\infty} \sqrt[k]{p_k} < 1$, one has the convolution identity*

$$b\big(\mathcal{R}(\boldsymbol{p})\big)_k = \frac{1}{k+1} \sum_{m=0}^{k} b(\boldsymbol{p})_m b(\boldsymbol{p})_{k-m}.$$

Proof. Let $\boldsymbol{p}$ be an arbitrary probability vector with $\limsup_{k\to\infty} \sqrt[k]{p_k} < 1$, so that the mapping b is well-defined. The left hand side leads to

$$\begin{aligned} b\big(\mathcal{R}(\boldsymbol{p})\big)_k &= \sum_{\ell\geq k} \binom{\ell}{k} \mathcal{R}(\boldsymbol{p})_\ell = \sum_{\ell\geq k} \binom{\ell}{k} \sum_{\substack{r,s\geq 0 \\ r+s\geq \ell}} \frac{p_r p_s}{r+s+1} \\ &= \sum_{\substack{r,s\geq 0 \\ r+s\geq k}} \frac{p_r p_s}{r+s+1} \sum_{\ell=k}^{r+s} \binom{\ell}{k} = \frac{1}{k+1} \sum_{\substack{r,s\geq 0 \\ r+s\geq k}} \binom{r+s}{k} p_r p_s, \end{aligned}$$

where a standard identity on binomial coefficients was used in the last step.

On the other hand, one finds

$$\begin{aligned} \sum_{m=0}^{k} b(\boldsymbol{p})_m b(\boldsymbol{p})_{k-m} &= \sum_{\substack{m,n\geq 0 \\ m+n=k}} \Big(\sum_{r\geq m} \binom{r}{m} p_r\Big)\Big(\sum_{s\geq n} \binom{s}{n} p_s\Big) \\ &= \sum_{\substack{m,n\geq 0 \\ m+n=k}} \sum_{\substack{r,s\geq 0 \\ r+s\geq k}} \binom{r}{m}\binom{s}{n} p_r p_s = \sum_{\substack{r,s\geq 0 \\ r+s\geq k}} p_r p_s \sum_{\substack{m,n\geq 0 \\ m+n=k}} \binom{r}{m}\binom{s}{n} \\ &= \sum_{\substack{r,s\geq 0 \\ r+s\geq k}} \binom{r+s}{k} p_r p_s, \end{aligned}$$

again using a standard identity, together with the fact that $\binom{n}{m} = 0$ for $m > n$ when n is an integer. A comparison of the two calculations establishes the claim. ∎

The further relevance of Lemma 3 stems from the following property of the generating function of $\boldsymbol{p}$, defined by $\psi(z) = \sum_{\ell\geq 0} p_\ell z^\ell$. When rewritten as a Taylor series around 1 rather than around 0, one obtains

$$\psi(z) = \sum_{\ell\geq 0} p_\ell z^\ell = \sum_{k\geq 0} \left(\sum_{\ell\geq k} \binom{\ell}{k} p_\ell \right) (z-1)^k = \sum_{k\geq 0} b(\boldsymbol{p})_k \, (z-1)^k. \tag{19}$$

Under the assumptions on $\boldsymbol{p}$, the radius of convergence of $\psi(z)$ is larger than 1, so that this calculation is on firm grounds. Lemma 3 now tells us that we may study the recombination action on the level of the expansion coefficients.

Let us therefore define the induced recombination operator $\widetilde{\mathcal{R}}$ on any space of type $X_{\alpha,\delta}$, with $\delta \geq \alpha$, by $\widetilde{\mathcal{R}}(b(\boldsymbol{a})) = b(\mathcal{R}(\boldsymbol{a}))$, which establishes a commuting diagram of the mappings $\mathcal{R}$ and $\widetilde{\mathcal{R}}$ in the obvious way. More precisely, one first restricts the action of $\mathcal{R}$ to a suitable subspace of $\mathcal{M}_1^+$, so that the mapping b is well-defined. If $\boldsymbol{p}$ satisfies the condition of Lemma 3, so that the radius of convergence of ψ exceeds 1, the probability vector $\boldsymbol{p}$ is also completely determined by its moments, compare [22, Thm. II.12.7] together with the observation that $\psi(e^{it})$ is the (convergent) moment generating function of $\boldsymbol{p}$. As all moments, in turn, are specified by the entries of $b(\boldsymbol{p})$, the latter uniquely determines $\boldsymbol{p}$ in this situation.

It is easy to check that the vector $(1, \alpha, \alpha^2, \ldots)$ is a fixed point of $\widetilde{\mathcal{R}}$ in $X_{\alpha,\delta}$, for any $\delta \geq \alpha$. Choosing $\alpha = \mathfrak{m}$, this vector is the image of the probability vector $\boldsymbol{p}$ from Proposition 4 under the mapping b.

Proposition 5. *On $X_{\alpha,\delta}$, the map defined by $\widetilde{\mathcal{R}}$ is a contraction. In particular, it is a globally Lipschitz continuous mapping of $X_{\alpha,\delta}$ into itself.*

Proof. Let $\delta \geq \alpha > 0$ be given, as well as arbitrary $\boldsymbol{a}, \boldsymbol{b} \in X_{\alpha,\delta}$. Clearly, we have $\widetilde{\mathcal{R}}(\boldsymbol{a})_0 = 1$ and $\widetilde{\mathcal{R}}(\boldsymbol{a})_1 = \alpha$. For $k \geq 2$, one finds $\widetilde{\mathcal{R}}(\boldsymbol{a})_k = \frac{1}{k+1} \sum_{\ell=0}^k a_\ell a_{k-\ell} \leq \delta^k$. This proves that $\widetilde{\mathcal{R}}$ maps $X_{\alpha,\delta}$ into itself.

The space $X_{\alpha,\delta}$ is equipped with the metric d from (17). Since, due to $\boldsymbol{b} \in X_{\alpha,\delta}$, also $\widetilde{\mathcal{R}}(\boldsymbol{b})_0 = 1$ and $\widetilde{\mathcal{R}}(\boldsymbol{b})_1 = \alpha$, the contraction estimate reads as follows.

$$\begin{aligned} d(\widetilde{\mathcal{R}}(\boldsymbol{a}), \widetilde{\mathcal{R}}(\boldsymbol{b})) &= \sum_{k\geq 2} \frac{d_k}{k+1} \Bigl| \sum_{\ell=0}^k (a_\ell a_{k-\ell} - b_\ell b_{k-\ell}) \Bigr| = \sum_{k\geq 2} \frac{d_k}{k+1} \Bigl| \sum_{\ell=0}^k (a_\ell - b_\ell)(a_{k-\ell} + b_{k-\ell}) \Bigr| \\ &\leq \sum_{k\geq 2} \frac{2\, d_k}{k+1} \sum_{\ell=2}^k \delta^{k-\ell} \, |a_\ell - b_\ell| = \sum_{\ell\geq 2} d_\ell \, |a_\ell - b_\ell| \sum_{k\geq \ell} \frac{2}{k+1} \, \delta^{k-\ell} \, \frac{d_k}{d_\ell}. \end{aligned}$$

With the choice $d_k = (\gamma/\delta)^k$, where we had $\gamma < \frac{1}{3}$, we can now find, for $\ell \geq 2$, an upper bound for the inner sum,

$$\sum_{k\geq \ell} \frac{2}{k+1} \, \delta^{k-\ell} \, \frac{d_k}{d_\ell} \leq \frac{2}{3} \sum_{k\geq \ell} \gamma^{k-\ell} = \frac{2}{3-3\gamma} =: C < 1,$$

which, together with the previous calculation, proves the contraction property,

$$d(\widetilde{\mathcal{R}}(\boldsymbol{a}), \widetilde{\mathcal{R}}(\boldsymbol{b})) \leq C \, d(\boldsymbol{a}, \boldsymbol{b}),$$

with contraction constant $C < 1$. Clearly, this also means that $\widetilde{\mathcal{R}}$ is globally Lipschitz continuous. ∎

This shows that, in discrete time, we have exponentially fast convergence of the sequence $(\widetilde{\mathcal{R}}^n(\boldsymbol{a}))_{n\geq 1}$, with $\boldsymbol{a} \in X_{\alpha,\delta}$, to a unique fixed point in $X_{\alpha,\delta}$. It is specified by the mean copy number $\mathfrak{m}$ of the probability vector $\boldsymbol{p}$ that underlies $\boldsymbol{a} = b(\boldsymbol{p})$, via $\alpha = \mathfrak{m}$, see above. Clearly, this fixed point (in $X_{\alpha,\delta}$) is the image (under b) of the equilibrium vector $\boldsymbol{p} \in \mathcal{M}_1^+$ calculated earlier in Proposition 4, as the mapping b is invertible in this situation. The claim of Theorem 1 for discrete time is now clear, with exponentially fast convergence to the equilibrium, from any initial condition as specified there.

For the slightly more involved treatment of the continuous time case, we refer to [19]. It is based on the construction of a Lyapunov function, similar to that of [19, Prop. 13].

6. Internal crossover. Another rather obvious model is based on the assumption that the shorter of the two sequences (or stretches) can align with any connected block of the longer sequence, but without any overhang. This situation has been coined internal unequal crossover, or *internal crossover* for short. Here, restricting to probability measures on $\mathbb{N}_0$, the recombinator (2) simplifies to

$$\mathcal{R}_0(\boldsymbol{p})_i = \sum_{\substack{k,\ell\geq 0 \\ k\wedge\ell\leq i\leq k\vee\ell}} \frac{p_k p_\ell}{1+|k-\ell|}, \tag{20}$$

where $k \wedge \ell$ ($k \vee \ell$) stands for the minimum (the maximum) of k and ℓ, see [23, 19, 18] for details on this model. We choose the notation $\mathcal{R}_0$ for reasons that will become clear later on.

In our search for fixed points, it is again useful to look for probability vectors that are reversible for $\mathcal{R}_0$. Since both forward and backward transition probabilities are simultaneously non-zero only when $\{i,j\} = \{k,\ell\} \subset \{n, n+1\}$ for some n, the components p_k may only be positive on this small set as well. By the following proposition, this indeed characterises all fixed points of this case.

Proposition 6. *A probability measure $\boldsymbol{p} \in \mathcal{M}_1^+$ is a fixed point of $\mathcal{R}_0$ if and only if its mean copy number $\mathfrak{m} = \sum_{k\geq 0} k\, p_k$ is finite, together with $p_{\lfloor \mathfrak{m}\rfloor} = \lfloor \mathfrak{m}\rfloor + 1 - \mathfrak{m}$, $p_{\lceil \mathfrak{m}\rceil} = \mathfrak{m} + 1 - \lceil \mathfrak{m}\rceil$, and $p_k = 0$ for all other k. This includes the case that $\mathfrak{m}$ is a non-negative integer, where $p_{\lfloor \mathfrak{m}\rfloor} = p_{\lceil \mathfrak{m}\rceil} = p_{\mathfrak{m}} = 1$.*

Proof. The 'if' follows easily by insertion into (13) and Lemma 2. For the 'only if' part, let i denote the smallest integer such that $p_i > 0$. Then,

$$\mathcal{R}(\boldsymbol{p})_i = p_i^2 + 2p_i \sum_{\ell\geq 1} \frac{p_{i+\ell}}{1+\ell} = p_i\Big(p_i + p_{i+1} + \sum_{\ell\geq 2} \frac{2}{\ell+1} p_{i+\ell}\Big) \leq p_i,$$

where the last step follows since $\frac{2}{\ell+1} < 1$ in the last sum. One has equality precisely when $p_k = 0$ for all $k \geq i+2$. This implies $\mathfrak{m} < \infty$ and the uniqueness of $\boldsymbol{p}$ (given $\mathfrak{m}$) with the non-zero frequencies as claimed. ∎

In this case, one may select a compact subset within the probability vectors by demanding the existence of the centred r-th moment, for some fixed $r > 1$. More precisely,

with

$$\mu_s(\boldsymbol{p}) := \sum_{\ell\geq 0} |\ell-\mathfrak{m}|^s p_\ell,$$

one considers the set

$$\mathcal{M}^+_{1,\mathfrak{m},C} := \{\boldsymbol{p}\in\mathcal{M}^+_1 \mid \textstyle\sum_k kp_k = \mathfrak{m} \text{ and } \mu_r(\boldsymbol{p})\leq C\} \tag{21}$$

for an arbitrary, but fixed $C<\infty$, equipped with our usual metric as introduced before in (1). This gives a compact and convex space [19, Lemma 2]. Moreover, one has

Lemma 4. *Let $r>1$ be fixed and consider the space $\mathcal{M}^+_{1,\mathfrak{m},C}$ of (21). Then, both μ_1 and μ_r satisfy*

$$\mu_s(\mathcal{R}_0(\boldsymbol{p}))\leq \mu_s(\boldsymbol{p}),$$

with equality if and only if $\boldsymbol{p}$ is a fixed point of $\mathcal{R}_0$.

Moreover, $\mu_1 : \mathcal{M}^+_{1,\mathfrak{m},C}\longrightarrow\mathbb{R}_{\geq 0}$ is continuous and defines a Lyapunov function for the dynamics in continuous time.

Proof. To show the first claim, consider

$$\begin{aligned}\mu_s(\mathcal{R}_0(\boldsymbol{p})) &= \sum_{i\geq 0}\sum_{\substack{k,\ell\geq 0\\ k\wedge\ell\leq i\leq k\vee\ell}} \frac{|i-\mathfrak{m}|^s}{1+|k-\ell|}\,p_k\,p_\ell\\ &= \sum_{k,\ell\geq 0}\frac{p_k\,p_\ell}{1+|k-\ell|}\,\frac{1}{2}\sum_{i=k\wedge\ell}^{k\vee\ell}(|i-\mathfrak{m}|^s+|k+\ell-i-\mathfrak{m}|^s).\end{aligned} \tag{22}$$

For notational convenience, let $j=k+\ell-i$. We now show

$$|i-\mathfrak{m}|^s+|k+\ell-i-\mathfrak{m}|^s\leq |k-\mathfrak{m}|^s+|\ell-\mathfrak{m}|^s. \tag{23}$$

If $\{k,\ell\}=\{i,j\}$, then (23) holds with equality. Otherwise, assume without loss of generality that $k<i\leq j<\ell$. If $\mathfrak{m}\leq k$ or $\mathfrak{m}\geq\ell$, we have equality for $s=1$, but a strict inequality for $s=r$ due to the convexity of $x\mapsto x^r$. (For $s=1$, this describes the fact that a recombination event between two sequences that are both longer or both shorter than the mean does not change their averaged distance to the mean copy number.) In the remaining cases, the inequality is strict as well. Hence, $\mu_s(\mathcal{R}_0(\boldsymbol{p}))\leq\mu_s(\boldsymbol{p})$ with equality if and only if $\boldsymbol{p}$ is a fixed point of $\mathcal{R}_0$, since otherwise the sum in (22) contains at least one term for which (23) holds as a strict inequality.

To see that μ_1 is continuous, consider a convergent sequence $(\boldsymbol{p}^{(n)})$ in $\mathcal{M}^+_{1,\mathfrak{m},C}$ and the random variables $H^{(n)}=|K^{(n)}-\mathfrak{m}|$, where the $K^{(n)}$ are independent $\mathbb{N}_0$-valued random variables with laws $\boldsymbol{p}^{(n)}$. Due to the structure of $\mathcal{M}^+_{1,\mathfrak{m},C}$, the random variables $H^{(n)}$ are uniformly integrable, which implies the convergence of the corresponding expectation values by [5, Thm. 25.12]. This, in turn, is nothing but the continuity of μ_1. Since $\mu_1(\boldsymbol{p})$ is linear in $\boldsymbol{p}$ and thus infinitely differentiable, so is the solution $\boldsymbol{p}(t)$ for every initial condition $\boldsymbol{p}_0\in\mathcal{M}^+_{1,\mathfrak{m},C}$, compare [1, Thm. 9.5 and Remark 9.6(b)]. Therefore, we have

$$\dot{\mu}_1(\boldsymbol{p}_0)=\liminf_{t\to 0^+}\frac{\mu_1(\boldsymbol{p}(t))-\mu_1(\boldsymbol{p}_0)}{t} \;=\; \mu_1(\mathcal{R}_0(\boldsymbol{p}_0))-\mu_1(\boldsymbol{p}_0)\leq 0,$$

again with equality if and only if $\boldsymbol{p}_0$ is a fixed point. Thus, μ_1 is a Lyapunov function as claimed. ∎

Finally, this gives the following convergence result, the proof of which is given in [19] and not repeated here.

THEOREM 2. *Assume that, for the initial condition $\boldsymbol{p}(0)$ and fixed $r > 1$, the r-th moment exists, $\mu_r(\boldsymbol{p}) < \infty$. Then, $\mathfrak{m} = \sum_\ell \ell p_\ell$ is finite and, both in discrete and in continuous time, $\lim_{t\to\infty} \|\boldsymbol{p}(t) - \boldsymbol{p}\|_1 = 0$ with the appropriate fixed point $\boldsymbol{p}$ from Proposition 6.* ∎

Let us mention that, for $q = 0$, the recombinator can be expressed in terms of explicit frequencies $\pi_{k,\ell}$ of fragment pairs before concatenation (with copy numbers k and ℓ) as $\mathcal{R}_0(\boldsymbol{p})_i = \sum_{j=0}^{i} \pi_{j,i-j}$. It is as yet an open question whether this can be used to simplify the above treatment.

7. Random crossover. This model deviates from the previous one in that it admits arbitrary overhangs, up to the case where, after the crossover, one sequence got it all while the other lost everything. The possible alignments for any pair are supposed to be equally likely, so that the recombinator (2), again restricted to the probability measures, now reads

$$\mathcal{R}_1(\boldsymbol{p})_i = \sum_{\substack{k,\ell\geq 0\\ k+\ell\geq i}} \frac{1+\min\{k,\ell,i,k+\ell-i\}}{(k+1)(\ell+1)}\, p_k\, p_\ell. \tag{24}$$

As for our previous two examples, using Lemma 2 once again, the reversibility condition,

$$\frac{p_k}{k+1}\frac{p_\ell}{\ell+1} = \frac{p_i}{i+1}\frac{p_j}{j+1}, \quad \text{for all } k+\ell = i+j,$$

leads to an expression for fixed points. In fact, these relations have $p_k = C(k+1)x^k$ as a solution, with appropriate parameter x and normalisation constant C. Again, it turns out that all fixed points are given this way, as was originally noticed (in a different way) in [23, Thm. A.2].

PROPOSITION 7. *Every fixed point $\boldsymbol{p} \in \mathcal{M}_1^+$ of $\mathcal{R}_1$ has finite mean $\mathfrak{m} = \sum_k k p_k$, and is uniquely specified by the value of $\mathfrak{m}$. Explicitly, one has*

$$p_k = \left(\frac{2}{\mathfrak{m}+2}\right)^2 (k+1)\left(\frac{\mathfrak{m}}{\mathfrak{m}+2}\right)^k$$

with $k \in \mathbb{N}_0$. ∎

One can verify this in several ways, one being a direct calculation via induction. Interestingly, this equilibrium is the convolution of two geometric distributions (of equal mean $\mathfrak{m}/2$), and hence also of geometric type according to our terminology (which follows that of [23]). It might be interesting to explore this observation a little further in the future.

At this point, one can define, very much in analogy to the situation in Takahata's model above, an induced recombinator, $\widetilde{\mathcal{R}}_1$, acting once more on spaces of the form $X_{\alpha,\delta}$. It is given as

$$\widetilde{\mathcal{R}}_1(\boldsymbol{p}) = a\big(\mathcal{R}_1(\boldsymbol{p})\big)$$

where the mapping a is defined by

$$a(\boldsymbol{p})_k = \frac{1}{k+1} \sum_{\ell \geq k} \binom{\ell}{k} p_\ell = \frac{1}{k+1}\, b(\boldsymbol{p})_k .$$

It is thus closely related to our above mapping b.

The main result on this model, proved in detail in [19, 18], reads as follows.

THEOREM 3. *Assume that* $\limsup_{k\to\infty} \sqrt[k]{p_k(0)} < 1$. *Then, both in discrete and in continuous time,* $\lim_{t\to\infty} \|\boldsymbol{p}(t) - \boldsymbol{p}\|_1 = 0$, *where* $\boldsymbol{p}$ *is the corresponding fixed point according to Proposition* 7.

Proof. The proof is very similar to the one used above for the Takahata model, and employs once again Banach's contraction principle for the induced action of $\widetilde{\mathcal{R}}_1$ on $X_{\alpha,\delta}$. Since all details have been given in [19], we omit them here. ∎

The fixed points of Proposition 7 are of the expected geometric type, and are perhaps more realistic than those of the Takahata model, at least for cases where a maximum is present in the repeat distribution. However, one should note that the experimental situation is not completely convincing at present, so that it seems advantageous to have a versatile model class at hand.

8. An interpolation. When considering the recombinators $\mathcal{R}_0$ and $\mathcal{R}_1$ in comparison, one would like to find further models that share properties of both of them, or interpolate between them in a suitable way. In particular, $\mathcal{R}_0$ is unrealistic due to the complete confinement of the shorter bit within the range of the longer one, while $\mathcal{R}_1$ poses no restriction at all for any kind of overhang. One such interpolation was initially investigated in discrete time by Atteson and Shpak in [23], based on preceding work by Ohta [17] and Walsh [26], see also [19, 18] for more. The interpolation employs a penalty function idea for overhangs of the shorter sequence, and leads (in the above language) to a recombinator $\mathcal{R}_q$ with $0 \leq q \leq 1$. The latter is based upon the transition probabilities

$$T^{(q)}_{ij,k\ell} = C^{(q)}_{k\ell}\, \delta_{i+j,k+\ell}\, (1 + \min\{k,\ell,i,j\})\, q^{0\vee(k\wedge\ell - i\wedge j)}, \tag{25}$$

where $k \vee \ell := \max\{k,\ell\}$, $k \wedge \ell := \min\{k,\ell\}$, and $0^0 = 1$. The normalisation constants $C^{(q)}_{k\ell}$ are chosen such that (3) holds, i.e., $\sum_{i,j\geq 0} T^{(q)}_{ij,k\ell} = 1$. These constants are symmetric in k and ℓ and read explicitly

$$C^{(q)}_{k\ell} = \frac{(1-q)^2}{(k\wedge\ell+1)(|k-\ell|+1)(1-q)^2 + 2q(k\wedge\ell - (k\wedge\ell+1)q + q^{k\wedge\ell+1})} .$$

Note further that the total number of units is indeed conserved in each event and that the process is symmetric within both pairs. Hence (7) is satisfied.

Unfortunately, the situation with the fixed points is a lot more complicated due to the following result.

PROPOSITION 8. *For parameter values* $q \in (0,1)$, *any fixed point* $\boldsymbol{p} \in \mathcal{M}_1^+$ *of the recombinator* $\mathcal{R}_q$, *given by* (2) *and* (25), *satisfies* $p_k > 0$ *for all* $k \geq 0$ (*unless it is the trivial fixed point* $\boldsymbol{p} = (1,0,0,\ldots)$ *we excluded*). *None of these extra fixed points is reversible.*

Proof. Let a non-trivial fixed point $\boldsymbol{p}$ be given and choose any $n > 0$ with $p_n > 0$. Observe that $T^{(q)}_{n+1\;n-1,nn} > 0$ for $0 < q < 1$ and hence

$$p_{n\pm1} = \mathcal{R}_q(\boldsymbol{p})_{n\pm1} = \sum_{j,k,\ell\geq 0} T^{(q)}_{n\pm1\;j,k\ell} p_k\, p_\ell \;\geq\; T^{(q)}_{n+1\;n-1,nn}\, p_n\, p_n > 0.$$

The first statement now follows by induction. For the second statement, evaluate the reversibility condition (13) for all combinations of i, j, k, ℓ with $i + j = k + \ell \leq 4$. This leads to four independent equations. Three of them can be transformed to the recursion

$$p_k = \frac{(k+1)q}{2(k-1)+2q}\,\frac{p_1}{p_0}\,p_{k-1}, \qquad k \in \{2,3,4\},$$

from which one derives explicit equations for all p_k with $k \in \{2,3,4\}$ in terms of p_0 and p_1. Inserting the one for p_2 into the remaining equation yields another equation for p_4 in terms of p_0 and p_1, which contradicts the first equation for all $q \in (0,1)$, as is easily verified. ∎

Nevertheless, the dynamics is well defined, and respects the compact subsets defined above in forward time, compare [19, Thm. 4]. Based upon the analysis in [18, 19], and further numerical work on the fixed points, it is plausible that, given the mean copy number $\mathfrak{m}$, never more than one fixed point for $\mathcal{R}_q$ exists. Due to the global convergence results at $q = 0$ and $q = 1$, any non-uniqueness in the vicinity of these parameter values could only come from a bifurcation, not from an independent source. Numerical investigations indicate that no bifurcation is present, but this needs to be analysed further.

Moreover, the Lipschitz constant for the corresponding induced recombinator $\widetilde{\mathcal{R}}_q$ can be expected to be continuous in the parameter q, hence to remain strictly less than 1 on the sets $X_{\alpha,\delta}$ in the neighbourhood of $q = 1$. So, at least locally, the contraction property should be preserved. For further progress, it seems advantageous [11] to use a rather different approach based on the analysis of similar problems in evolutionary game theory. Here, one would aim to establish a slightly weaker type of convergence result for all $0 < q < 1$, and probably even on the larger compact set $\mathcal{M}^{+}_{1,\mathfrak{m},C}$ from Eq. (21).

9. Open problems and outlook. The results for the various models presented here show that initial configurations, subject to some specific conditions that are no restriction in practice, converge to one of the known fixed points. These results apply to the deterministic dynamics of the infinite population limit.

In view of the biological applications, one is also interested in possible deviations from this picture on the level of large, but finite, populations, i.e., for the underlying stochastic process, e.g., a variant of the Moran model with unequal crossover. In this model class, however, important deviations seem unlikely, due to the known convergence results for the infinite population limit, see [2] for more.

Since the above equilibrium distributions have finite support or are exponentially small for large copy numbers, one can also expect these systems to behave very much like ones with only finitely many types. In this sense, the results are typical, and the more general setting with probability vectors on $\mathbb{N}_0$ is adequate. This is also supported by several simulations [18].

Still, an open question is a more complete understanding of the regime $q \in (0,1)$ in Section 8. Due to the loss of reversibility of the fixed points, the analysis becomes rather involved. Preliminary investigations [18] have not given any hint on values of q where convergence fails or where alternative stable fixed points show up, though this is presently only based on numerical experiments and perturbative arguments. It might be advantageous (and perhaps also more realistic) to search for other ways to interpolate between the cases $q = 0$ and $q = 1$, preferably ones that maintain the reversibility of the equilibria. This question certainly deserves further attention.

References

[1] H. Amann, *Ordinary Differential Equations*, de Gruyter, Berlin, 1990.

[2] E. Baake and I. Hildebrandt, *Single-crossover dynamics: finite versus infinite populations*, Bull. Math. Biol. 70 (2008), 603–624.

[3] D. Bachtrog, S. Weiss, B. Zangerl, G. Brem and C. Schlötterer, *Distribution of dinucleotide microsatellites in the* Drosophila melanogaster *genome*, Mol. Biol. Evol. 16 (1999), 602–610.

[4] C. Bank, *Diskrete Rekombinationsdynamik für repetitive Strukturen*, Diplomarbeit, Univ. Bielefeld, 2007.

[5] P. Billingsley, *Probability and Measure*, 3rd ed., Wiley, New York, 1995.

[6] P. Billingsley, *Convergence of Probability Measures*, 2nd ed., Wiley, New York, 1999.

[7] R. Bürger, *The Mathematical Theory of Selection, Recombination and Mutation*, Wiley, Chichester, 2000.

[8] C. D. Bustamante, *Population genetics of molecular evolution*, in [16], pp. 63–99.

[9] P. Calabrese and R. Sainudiin, *Models of microsatellite evolution*, in [16], pp. 289–305.

[10] D. Graur and W.-H. Li, *Fundamentals of Molecular Evolution*, 2nd ed., Sinauer, Sunderland, 2000.

[11] J. Hofbauer, private communication, 2003.

[12] B. Kitchens, *Symbolic Dynamics—One-sided, Two-sided and Countable State Markov Chains*, Springer, Berlin, 1998.

[13] S. Kruglyak, R. T. Durrett, M. D. Schug and C. F. Aquadro, *Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations*, Proc. Natl. Acad. Sci. USA 95 (1998), 10774–10778.

[14] M. Lachowicz, *General population systems. Macroscopic limit of a class of stochastic semigroups*, J. Math. Anal. Appl. 307 (2005), 585–605.

[15] M. Lachowicz, *Micro and meso scales of description corresponding to a model of tissue invasion by solid tumors*, Math. Models Meth. Appl. Sci. 15 (2005), 1667–1683.

[16] R. Nielsen (ed), *Statistical Methods in Molecular Evolution*, Springer, New York, 2005.

[17] T. Ohta, *On the evolution of multigene families*, Theor. Pop. Biol. 23 (1983), 216–240.

[18] O. Redner, *Models for mutation, selection, and recombination in infinite populations*, Dissertation, Univ. Greifswald; Shaker, Aachen, 2003.

[19] O. Redner and M. Baake, *Unequal crossover dynamics in discrete and continuous time*, J. Math. Biol. 49 (2004), 201–226; arXiv:math.DS/0402351.

[20] M. Reed and B. Simon, *Methods of Modern Mathematical Physics I: Functional Analysis*, Academic Press, San Diego, 1980.

[21] E. Seneta, *Non-negative Matrices and Markov Chains*, rev. printing, Springer, New York, 2006.
[22] A. N. Shiryaev, *Probability*, 2nd ed., Springer, New York, 1996.
[23] M. Shpak and K. Atteson, *A survey of unequal crossover systems and their mathematical properties*, Bull. Math. Biol. 64 (2002), 703–746.
[24] N. Takahata, *A mathematical study on the distribution of the number of repeated genes per chromosome*, Genet. Res. 38 (1981), 97–102.
[25] J. Wakeley, *Coalescent Theory: An Introduction*, Roberts and Company, Greenwood Village, CO, 2008.
[26] J. B. Walsh, *Persistence of tandem arrays: Implications for satellite and simple-sequence DNAs*, Genetics 115 (1987), 553–567.
[27] K. Yosida, *Functional Analysis*, 6th ed., Springer, Berlin, 1980.

STOCHASTIC MODELS IN BIOLOGICAL SCIENCES
BANACH CENTER PUBLICATIONS, VOLUME 80
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WARSZAWA 2008

DETERMINISTIC AND STOCHASTIC SIMULATIONS OF SIMPLE GENETIC CIRCUITS

OFER BIHAM, NATHALIE Q. BALABAN, ADIEL LOINGER

Racah Institute of Physics, The Hebrew University, Jerusalem 91904, Israel
E-mail: biham@phys.huji.ac.il, nathalieqb@phys.huji.ac.il, adiel@phys.huji.ac.il

AZI LIPSHTAT

Department of Pharmacology and Biological Chemistry, Mount Sinai School of Medicine, New York, NY 10029, U.S.A.
E-mail: azi.lipshtat@mssm.edu

HAGAI B. PERETS

Faculty of Physics, Weizmann Institute of Science, Rehovot 76100, Israel
E-mail: hagai.perets@weizmann.ac.il

Abstract. We analyze three simple genetic circuits which involve transcriptional regulation and feedback: the autorepressor, the switch and the repressilator, that consist of one, two and three genes, respectively. Such systems are commonly simulated using rate equations, that account for the concentrations of the mRNAs and proteins produced by these genes. Rate equations are suitable when the concentrations of the relevant molecules in a cell are large and fluctuations are negligible. However, when some of the proteins in the circuit appear in low copy numbers, fluctuations become important and the rate equations fail. In this case stochastic methods, such as direct numerical integration of the master equation or Monte Carlo simulations are required. Here we present deterministic and stochastic simulations of the autorepressor, the switch and the repressilator. We show that fluctuations give rise to quantitative and qualitative changes in the dynamics of these systems. In particular, we demonstrate a fluctuations-induced bistability in a variant of the genetic switch and and noisy oscillations obtained in the repressilator circuit.

2000 *Mathematics Subject Classification*: Primary 92D10.

Key words and phrases: systems biology, genetic networks, genetic switch, feedback loop, stochastic analysis.

This work was supported by the Center for Complexity Studies.

The paper is in final form and no version of it will be published elsewhere.

 © Instytut Matematyczny PAN, 2008

1. Introduction. The production of proteins in cells is regulated by networks of interacting genes. Some of these genes code for transcription factors, which are proteins that regulate the expression of other genes by binding to specific promoter sites on the DNA. Some of these proteins, called repressors, perform negative regulation, while others, called activators, perform positive regulation. Post-trancriptional regulation can occur by translational regulation, or by post-translational regulation such as protein-protein interaction, which may modify the function of these proteins. These networks of interacting genes are at the heart of all processes in cells: from the regulation of the cell cycle to the various stress responses.

It turns out that genetic networks are typically sparse networks, namely most genes interact only with a small number of other genes. Also, the networks exhibit some degree of modularity, namely one can identify recurring modules, which are known as network motifs [18].

Modules found in nature are often hard to study and fully control. To overcome these limitations, synthetic networks can be constructed from well-studied components. Examples of such components are the lactose, the lambda and the tetracycline systems in bacteria. The synthetic networks are designed to perform desired functions, determined by their architecture. They do not require the manipulation of the structure of proteins and other regulatory elements at the molecular level. The genes and promoters are often inserted into plasmids rather than on the chromosome. Two important examples of synthetic circuits are the genetic toggle switch and the repressilator.

The toggle switch consists of two genes, which negatively regulate each other expression. The regulation is performed at the transcriptional level, namely each gene codes for a repressor protein that binds to the promoter of the other gene. Such system may exhibit bistability, namely two stable states, where in each state one of the proteins is dominant and the other is suppressed. A synthetic toggle switch was constructed in *E. coli* and the conditions for bistability were examined [6]. The switching between its two states was demonstrated using chemical and thermal induction. More recently, such circuit was found to exist in a natural system in which two mutual repressors regulate the differentiation of myeloid progenitors into either macrophages or neutrophils [11].

The synthetic repressilator circuit, constructed in *E. coli*, was designed to exhibit oscillations, reminiscent of natural genetic oscillators such as the circadian rhythms [4]. The repressilator circuit was encoded on a plasmid that appears in a low copy number. The protein concentrations were measured vs. time in single cells. Oscillations with a period of about 150 minutes were found, namely extending over several division cycles. The oscillations were found to be noisy, typically maintaining phase coherence for times of the order of a single oscillation period.

The dynamics of genetic networks is often simulated using rate equation models. These are sets of coupled ordinary differential equations, which account for the concentrations of the mRNAs and proteins in the network. In general, rate equations account for average concentrations and ignore fluctuations. They are suitable for systems in which the concentration of interacting molecules are large and fluctuations are negligible. However, proteins in cells often appear in low copy numbers and may exhibit large fluctuations.

Moreover, in case of transcriptional regulation, the expression of the regulated gene may be controlled by a single protein that binds to its promoter. This extends the notion of low copy numbers and the resulting fluctuations even to the case when there is a large number of free repressors of a certain type, since only one of them may bind to the promoter at any given time. Recent advances measurement techniques made it possible to measure the fluctuations in copy number of proteins in single cells [5, 23, 26]. Measurements of protein levels in single cells revealed distributions that depend on the topology of the regulatory network controlling the particular protein [19].

To account for the fluctuations, simulations of genetic networks should be done using stochastic methods such as the master equation [10, 16, 17, 20, 21, 22, 23] or Monte Carlo simulations [7, 9, 16]. The master equation provides the probability distribution of the concentrations of proteins in a population of cells. From this distribution one can calculate the average concentrations as well as the correlations and the rates of all processes. Monte Carlo simulations enable to follow the temporal variations in the level of gene expression and in the concentrations of proteins in a single cell. These results enable to extract the noise level and temporal correlation functions. They can also be used to characterize the dynamics of the system and determine whether it exhibits a single steady state, multi-stability or oscillatory behavior.

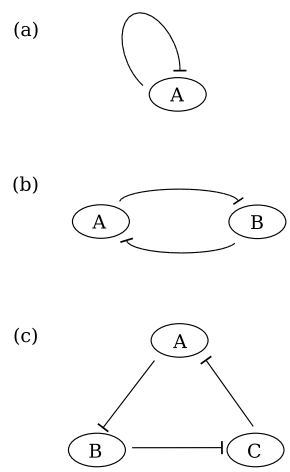


Fig. 1. The three genetic circuits considered in this paper: the autorepressor (a); the genetic switch (b); and the repressillator (c). The flat-headed arrows denote negative transcriptional regulation.

In this paper we present deterministic analysis (using rate equations) and stochastic analysis (using direct numerical integration of the master equation and Monte Carlo simulations) of three simple genetic circuits: the autorepressor, the toggle switch and the repressilator (Fig. 1). We show that fluctuations give rise to both quantitative and qualitative effect in the dynamics of thse circuits.

The paper is organized as follows. In Sec. 2 we consider the autorepressor and present the deterministic and stochastic methods used in the paper. In Sec. 3 we study the genetic switch and in Sec. 4 we analyze the repressillator circuit. The results are discussed and summarized in Sec. 5.

2. The autorepressor. The autorepressor is the simplest genetic circuit that involves feedback. It consist of a single gene, denoted by a, that negatively regulates its own expression. The gene transcribes into mRNAs that translate into A proteins. These proteins function as repressors. When an A protein binds to the a promoter site, it prevents the RNA polymerase from binding to the promoter and thus inhibits the transcription process. It turns out that genetic networks include a large number of autorepressor modules. This circuit is thus considered as a network motif [25]. One may speculate that it performs some crucial function in the cell. It was proposed that the role of the autorepressor is to speed up response times [25] or to reduce fluctuations [2]. Below we analyze the autorepressor using deterministic and stochastic methods. We utilize its simplicity in order to present the methodologies in details.

2.1. *Michaelis-Menten equations.* The dynamics of genetic networks are commonly described using the Michaelis-Menten equations. These equations describe the temporal variations in the concentrations of the relevant molecules in the cell. Here, we denote the concentration of protein A in a cell by $[A]$ (by concentration we refer to the average copy number of A proteins per cell). The concentration of the corresponding mRNA is denoted by $[m]$. The Michaelis-Menten equations for the concentrations take the form

$$\begin{aligned} [\dot{m}] &= \frac{g_m}{1+k[A]^n} - d_m[m], \\ [\dot{A}] &= g_p[m] - d_p[A]. \end{aligned} \tag{1}$$

The parameters d_m and d_p ($\sec^{-1}$) are the degradation rates of the mRNAs and proteins, respectively. The trascription rate is given by g_m ($\sec^{-1}$). The rate in which each mRNA is translated into proteins is given by g_p ($\sec^{-1}$). Since A proteins negatively regulate their own synthesis, the transcription rate is reduced by a factor of $1/(1+k[A]^n)$. This factor is called the Hill-function. In this expression, the parameter k quantifies the regulation strength (determined by the affinity between the repressor and the promoter site). The exponent n is called the Hill-coefficient. In general, Hill-function models can be derived from more complete rate equation models. In this case, n is expected to take only integer values. In fact, n represents the number of copies of the transcription factor, that are required to be bound simultaneously to the promoter in order to perform the regulation. The case of $n > 1$ is often referred to as cooperative binding. In this paper we consider only integer values of n. However, similar models may also be used to fit empirical data. In this case, n is simply a fitting parameter which may take non-integer values.

To simplify the analysis of genetic circuits, the mRNA level is often ignored and the transcription and translation processes are regarded as a single step of protein synthesis [25]. In this case, the effective production rate of proteins is given by $g = g_p g_m/d_m$. The

Michaelis-Menten equations are reduced to

$$\dot{[A]} = \frac{g}{1 + k[A]^n} - d[A], \tag{2}$$

where $d = d_p$. Ignoring the mRNA level is typically justified under steady state conditions. However, in the analysis of systems that are away from steady state due to external signals, or those that exhibit oscillations or large fluctuations the mRNA level should be included. The Michaelis-Menten equations for the autorepressor exhibit a single steady state solution for the concentration of A proteins. In case that $n = 1$, it is given by

$$[A] = \frac{-1 + \sqrt{1 + 4kg/d}}{2k}. \tag{3}$$

This solution is found to be stable for any choice of the parameters.

2.2. *Extended set of rate equations.* In the Michaelis-Menten equations the negative regulation process is described by the Hill-function. This description is rather crude and incomplete. In order to model the regulation process in greater detail we present below a more complete rate equation model [12]. In this model, we account separably for the populations of free and bound proteins. The bound A proteins are denoted by r and their concentration is given by $[r]$.

Consider an autorepressor gene a, encoded on the chromosome, which exhibits no cooperative binding. In this case the number of bound repressors is in the range $0 \leq [r] \leq 1$. The concentration, $[r]$, can also be considered as the fraction of time in which the promoter is occupied by a bound repressor and the transcription process is suppressed. Therefore the transcription rate is reduced by a factor of $(1 - [r])$. Ignoring the mRNA level, the extended set of rate equations takes the form [12]

$$\begin{aligned} \dot{[A]} &= g(1 - [r]) - d[A] - \alpha_0[A]\,(1 - [r]) + \alpha_1[r], \\ \dot{[r]} &= \alpha_0[A]\,(1 - [r]) - \alpha_1[r], \end{aligned} \tag{4}$$

where the parameter α_0 ($\sec^{-1}$) is the binding rate of the repressors to the promoter site and α_1 ($\sec^{-1}$) is the their unbinding rate. In the limit in which the binding and unbinding processes are much faster than other processes in the system (namely $\alpha_0, \alpha_1 \gg d, g$), these equations can be reduced to the Michaelis-Menten form. In this limit, the relaxation time of the concentration $[r]$ is much shorter than other relaxation times in the system. Therefore, one can take the time derivative of $[r]$ to zero, even if the system is away from steady state. This brings the rate equations to the Michaelis-Menten form [Eq. (2)] with $n = 1$ and $k = \alpha_0/\alpha_1$. Therefore, Eqs. (4) have the same steady state solution for the protein A as the Michaelis-Menten equation (2). However, the dynamics leading to steady state may differ between the two equations. Furthermore, the extended rate equation model exhibits more flexibility in the sense that it is much easier to insert additional features into Eqs. (4) than into Eq. (2). For example, it is possible to introduce degradation of bound repressors by adding the term $-d_r[r]$ to the equation for $\dot{[r]}$ in Eqs. (4).

2.3. *Stochastic analysis.* Transcription factors and other proteins, as well as their mRNAs in a cell, often appear in low concentrations [9, 17, 8]. In this case, fluctuations in the

copy numbers of these molecules may play an important role in the dynamics of genetic networks. To obtain a better description of thse systems, one should take into account the discrete nature of the molecules rather than using continuous concentrations. Moreover, even in case that some transcription factor appears in a high concentration, the regulation is performed by a small number of copies that are bound to the promoter site. The fluctuations in the number of bound transcription factors give rise to large temporal variations in the transcription rate of the regulated gene.

In order to account for fluctuations in genetic networks, stochastic methods are required, such as the master equation or Monte Carlo simulations [1, 10, 16, 20, 23]. The master equation is expressed in terms of the probability distribution $P(N_A, N_r)$. This is the probability for a cell to include $N_A = 0, 1, 2, \ldots$ free copies of protein A and $N_r = 0, 1$ copies of the same protein, which are bound to the promoter. The master equation accounts for the temporal variations in the probability distribution. For the autorepressor, it takes the form [12]

$$\begin{aligned}\dot{P}(N_A, N_r) = {} & g\delta_{N_r,0}[P(N_A - 1, N_r) - P(N_A, N_r)] \\ & + d[(N_A + 1)P(N_A + 1, N_r) - N_A P(N_A, N_r)] \\ & + \alpha_0[\delta_{N_r,1}(N_A + 1)P(N_A + 1, N_r - 1) - \delta_{N_r,0}N_A P(N_A, N_r)] \\ & + \alpha_1[\delta_{N_r,0}P(N_A - 1, N_r + 1) - \delta_{N_r,1}P(N_A, N_r)],\end{aligned} \tag{5}$$

where the g term accounts for the production of proteins and the d term accounts for their degradation. The α_0 (α_1) terms describe the binding (unbinding) of proteins to (from) the promoter site.

In numerical integration, the master equation must be truncated in order to keep the number of equations finite. This is done by setting a suitable upper cutoff, $N_{\mathrm{A}}^{\max}$, on the population size of the free proteins. In order to maintain the accuracy of the calculations, the cutoff should be chosen such that the probability of population sizes beyond it will be sufficiently small. The master equation exhibits a single steady state solution, characterized by $\dot{P}(N_A, N_r) = 0$ for all N_A and N_r. This solution is always stable [27]. The average concentration of free A proteins can be obtained from

$$\langle N_{\mathrm{A}} \rangle = \sum_{N_{\mathrm{A}}=0}^{N_{\mathrm{A}}^{\max}} \sum_{N_{\mathrm{r}}=0}^{1} N_{\mathrm{A}} P(N_A, N_r). \tag{6}$$

Other properties of the distribution, such as the variance, can be obtained from the calculation of higher moments.

Another useful approach to the study of stochastic dynamics is provided by Monte Carlo methods [7, 9, 16]. The simulation dynamics is Markovian. At each instant, the next process to take place is chosen randomly from all the possible processes, where each process is assigned with a suitable weight, proportional to its rate. The elapsed time is then updated accordingly. Unlike the master equation, which accounts for the entire distribution, Monte Carlo simulations follow the temporal variations in protein concentrations in a single cell.

In Fig. 2(a) we present the concentration of A proteins vs. time, obtained from the rate equations [Eq. (4)] and from Monte Carlo simulations. In Fig. 2(b) we show the

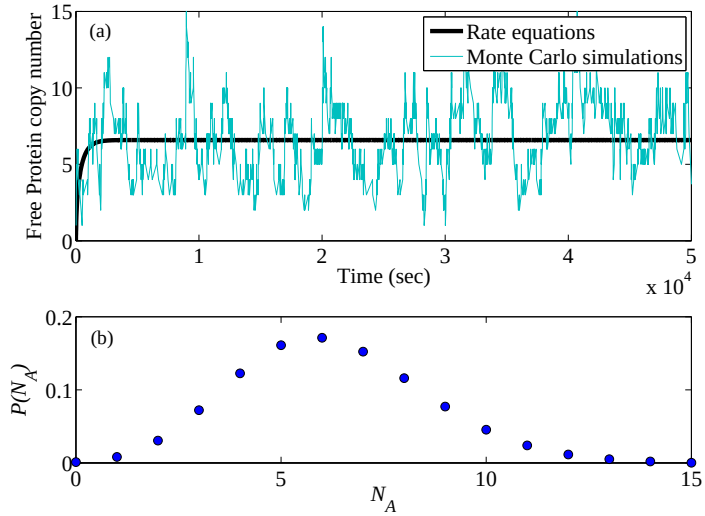


Fig. 2. Results for the autorepressor circuit. (a) The concentration of A proteins vs. time, obtained from the rate equations and from Monte Carlo simulations. The rate equation results quickly reach a steady state. The Monte Carlo results fluctuate around this steady state. (b) The steady state probability distribution $P(N_A)$ for a cell to contain N_A copies of protein A, obtained from the master equation. The parameters used are $g = 0.05$, $d = 0.001$, $\alpha_0 = 0.01$, $\alpha_1 = 0.01$ and $d_r = 0$ ($\sec^{-1}$).

probability distribution $P(N_{\rm A}$ obtained from the master equation under steady state conditions.

3. The genetic switch. The genetic toggle switch consists of two proteins, A and B, which negatively regulate each other at the transcriptional level [Fig. 1(b)]. This architecture may lead to two steady states, one dominated by A proteins and the other dominated by B proteins. When the population of A proteins is much larger than that of B proteins, the A proteins tend to suppress the production of B proteins. Under these conditions, the production of A proteins is enhanced, because the declining concentration of B proteins is not sufficient to suppress it. Therefore, the system approaches a state rich in A proteins and poor in B proteins. Similarly, the system may approach a state rich in B proteins and poor in A proteins. Transitions between the two states may take place in response to a suitable external signal. Spontaneous transitions, due to random fluctuations, are also possible. To qualify as a switch, the system should exhibit bistability. In the deterministic description, bistability is defined as the existence of two stable steady state solutions of the rate equations. This description does not account for the possibility of spontaneous transitions between the two states. In the stochastic description, spontaneous transitions are taken into account. Therefore, the condition for bistability is that the rate of spontaneous transitions (due to random fluctuations rather than an external signal) is much lower than the rates of all other relevant processes in the system.

Genetic switch systems exist in nature, and give rise to different cell behaviors in different situations. A notable example is the phage λ switch [24]. This switch appears in λ phages, which infect *E. coli* bacteria and can exist in two exclusive states, one called lysogeny and the other called lysis. In addition to natural switches, a synthetic switch was constructed and studied in *E. coli* [6]. Numerous studies, using rate equations, have concluded that cooperative binding is a necessary condition for the emergence of bistability [3, 6, 28, 29, 30]. Stochastic analysis reveals the reason to this fact. For a switch without cooperative binding, three peaks are obtained in the probability distribution function. These peaks corresponds to three possible states for the system: one in which A is highly expressed, a second in which B is highly expressed and a third in which both proteins are suppressed (a 'deadlock' situation) [13, 14]. Monte Carlo simulations show rapid transitions between these three states. The possibility of simultaneous suppression of both proteins, prevents the system from functioning as a switch. It causes the system to exhibit three states of limited stability instead of the two stable states that are desired.

It is found that in switch systems in which the A and B repressors exhibit cooperative binding, the deadlock situation is removed and bistability emerges. This can be explained as follows. The deadlock situation results from a simultaneous binding of A and B repressors to the corresponding promoter sites. Without cooperative binding, it is sufficient for the minority protein to recruit a single copy that will bind and suppress the production of the dominant protein. In the case of cooperative binding (for example, with $n = 2$) the minority protein needs to recruit two copies that will bind simultaneously in order to suppress the production of the dominant protein. This is much less likely. Therefore, cooperative binding induces bistability and enables the system to function as a switch.

Apart from cooperative binding, there are several other mechanisms that may stabilize bistability in genetic switch systems. The most obvious is the exclusive switch, where there is an overlap between the promoters of A and B, leaving no room for both to be occupied simultaneously. Such situations are encountered in nature, for example, in the lysis-lysogeny switch of phage λ [24]. It was shown that in the presence of cooperative binding, the exclusive binding of A and B repressors enhances the stability of the genetic switch [29]. This is because in exclusive binding the access of the minority specie to the promoter site is blocked by the dominant specie. For the exclusive switch without cooperative binding, rate equations exhibit only a single steady state solution, namely, the system is not a switch in the deterministic framework. However, in the stochastic framework the system exhibits bistability and functions as a switch. The distribution $P(N_A, N_B)$ exhibits two peaks, one dominated by A proteins and the other dominated by B proteins. The exclusive binding prevents the possibility of a deadlock situation.

In addition to the exclusive switch, there exist other variants of the genetic switch circuit (we focus here on systems without cooperative binding). Consider a switch in which not only free proteins, but also bound proteins experience degradation. Bound-repressor degradation tends to prevent the deadlock situation in which both A and B repressors are bound simultaneously. This is due to the fact that degradation removes the bound repressor from the system, unlike unbinding, where the resulting free repressor may quickly bind again. It turns out that degradation of bound repressors induces bistability not only in the stochastic framework but also at the level of deterministic rate equations.

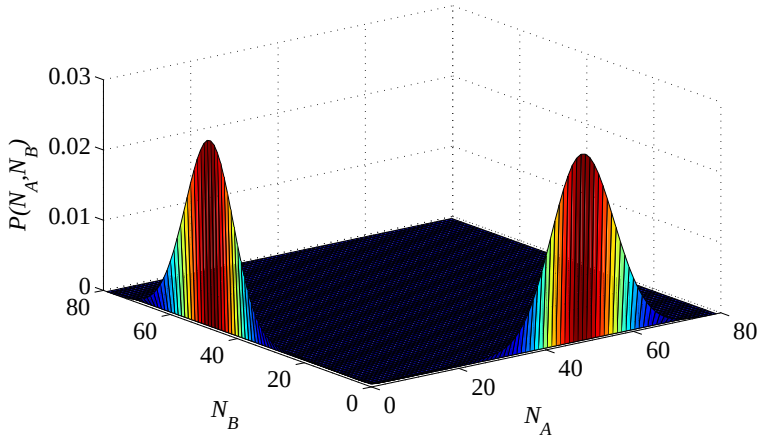


Fig. 3. The probability distribution $P(N_\mathrm{A}, N_\mathrm{B})$ of the concentrations of A and B proteins, for the switch with degradation of bound repressors, obtained from the master equation. Two sharp peaks are observed, one dominated by A proteins and the other dominated by B proteins. The peaks are sharp and are separated by a region with vanishing probabilities. As a result, the transition rate between the two peaks is low and the switch is stable.

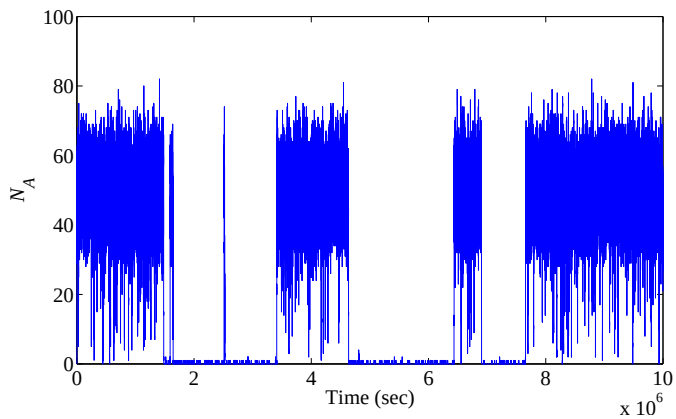


Fig. 4. The concentration, N_A, of A proteins vs. time, obtained from Monte Carlo simulations for the switch with degradation of bound repressors. The two states are clearly observed: one in which N_A fluctuates around 50, and another in which it is nearly zero. Transitions between these states occur at an average rate of one transition every $\sim 10^6$ sec (about 10 days).

In Fig. 3 we present the probability distribution $P(N_\mathrm{A}, N_\mathrm{B})$ of the concentrations of A and B proteins, for the switch with degradation of bound repressors. Two sharp peaks, well separated from each other are observed, illuminating the bistable nature of the systems. The parameters used in Fig. 3 and in the rest of the paper are: $g = 0.15$, $d = 0.003$, $\alpha_0 = 0.5$, $\alpha_1 = 0.01$ and $d_r = 0.003$ (sec^{-1}).

In Fig. 4 we present the temporal variations of the concentration, N_A, of A proteins, obtained from Monte Carlo simulations for the switch with degradation of bound repressors. Two states are clearly observed: one in which N_A is dominant and another state

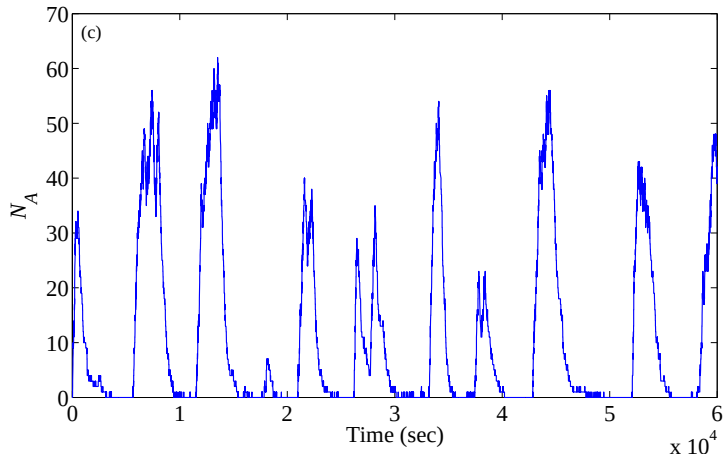


Fig. 5. The concentration, N_A, of A proteins vs. time, obtained from Monte Carlo simulations of the repressilator circuit. The oscillations are noisy. Their period and amplitude vary from cycle to cycle.

in which it is suppressed. Note that in spite of the very large fluctuations, the switch is stable and the average time between spontaneous transitions is about 10 days.

Another variant of the genetic switch, which exhibits bistability even at the level of rate equations, involves protein-protein interactions, where A and B proteins bind to each other and form a complex that does not function as a transcription factor. This additional process contributes to the stability of the switch because in such 'pair annihilation' processes the minority protein is affected more strongly. It is thus less likely to bind and suppress the production of the dominant protein.

4. The repressilator. The repressilator circuit consists of three transcription factors, A, B and C, which negatively regulate each other's synthesis in a cyclic manner [Fig. 1(c)]. This circuit was synthetically constructed on plasmids in *E. coli* and was found to exhibit oscillations in the concentrations of the three transcription factors. To understand the origin of these oscillations, consider a situation in which the number of A proteins is large. In this case it is likely that one of the A proteins will bind the to b promoter and will repress the production of B proteins. The reduced level of B proteins will enable the gene c to be fully expressed and the number of C proteins will increase and will start to repress gene a. As a result, the number of A proteins will decrease, and gene b will be activated, completing a full cycle. The order of appearance of the dominant protein type in this cycle is $A \to C \to B \to A$.

From a theoretical point of view, oscillations in this system can be obtained (under some conditions) both in rate equations and in Monte Carlo simulations [Fig. 5]. The oscillations obtained from the rate equations are regular, forming a stable limit-cycle. The oscillations obtained from the Monte Carlo simulations are noisy and irregular. Moreover, the period and amplitude differ significantly between the rate equations and the Monte Carlo simulations [15].

The repressilator system was constructed synthetically on plasmids. When the number of plasmids in a cell is small, the numbers of promoter sites and bound transcription factors are also small. As a result, one expects large fluctuations in the transcription rates regulated by these transcription factors. In this case stochastic methods are required. However, when the number of plasmids is large, fluctuations are reduced and the rate equations become applicable. Therefore, by gradually changing the number of plasmids in the cell one can explore the cross-over from the stochastic regime, where the oscillations are noisy, to the deterministic regime, where the oscillations are regular.

5. Summary. We have presented deterministic and stochastic analysis of three simple genetic circuits which involve transcriptional regulation and feedback: the autorepressor, the switch and the repressilator. Such systems can be simulated using rate equations, that account for the concentrations of the mRNAs and proteins produced by these genes. Rate equations are suitable when these concentrations are large and fluctuations are negligible. However, when some of the transcription factors and their binding sites in a cell appear in low copy numbers, fluctuations become important and the rate equations fail. In this case stochastic methods such as the master equation or Monte Carlo simulations are required. We have shown that fluctuations give rise to quantitative and qualitative changes in the dynamics of the systems. In particular, we demonstrated the fluctuations-induced bistability in the exclusive switch and the noisy oscillations obtained in the repressilator circuit.

References

[1] A. Arkin, J. Ross and H. H. McAdams, *Stochastic kinetic analysis of developmental pathway bifurcation in λ-infected E. coli cells*, Genetics 149 (1998), 1633.

[2] A. Becskei and L. Serrano, *Engineering stability in gene networks by autoregulation*, Nature 405 (2000), 590.

[3] J. L. Cherry and F. R. Adler, *How to make a biological switch*, J. Theor. Biol. 203 (2000), 117.

[4] M. B. Elowitz and S. Leibler, *A synthetic oscillatory network of transcriptional regulators*, Nature 403 (2000), 335.

[5] M. B. Elowitz, A. J. Levine, E. D. Siggia and P. S. Swain, *Stochastic gene expression in a single cell*, Science 297 (2002), 1183.

[6] T. S. Gardner, C. R. Cantor and J. J. Collins, *Construction of a genetic toggle switch in E. coli*, Nature 403 (2000), 339.

[7] M. A. Gibson and J. Bruck, *Efficient exact stochastic simulation of chemical systems with many species and many channels*, J. Phys. Chem. 104 (2000), 1876.

[8] M. A. Gibson and E. Mjolsness, *Modeling the activity of single genes*, in: J. M. Bower and H. Bolouri, (eds.), Computational Modeling of Genetic and Biochemical Networks, MIT Press, Cambridge, MA, 2000, 1–48.

[9] D. T. Gillespie, *Exact stochastic simulation of coupled chemical reactions*, J. Phys. Chem. 81 (1977), 2340.

[10] T. B. Kepler and T. C. Elston, *Stochasticity in transcriptional regulation: origins, consequences, and mathematical modeling representations*, Biophysical Journal 81 (2001), 3116.

[11] P. Laslo et al., *Multilineage transcriptional priming and determination of alternate hematopoetic cell fates*, Cell 126 (2006), 755.

[12] A. Lipshtat, H. B. Perets, N. Q. Balaban and O. Biham, *Modeling of negative autoregulated genetic networks in single cells*, Gene 347 (2005), 265.

[13] A. Lipshtat, A. Loinger, N.Q. Balaban and O. Biham, *Genetic toggle switch without cooperative binding*, Phys. Rev. Lett. 96 (2006), 188101.

[14] A. Loinger, A. Lipshtat, N. Q. Balaban and O. Biham, *Stochastic simulations of genetic switch systems*, Phys. Rev. E 75 (2007), 021904.

[15] A. Loinger and O. Biham, *Stochastic simulations of the repressilator circuit*, submitted to Phys. Rev. E (2007).

[16] H.H. McAdams and A. Arkin, *Stochastic mechanisms in gene expression*, Proc. Natl. Acad. Sci. USA 94 (1997), 814.

[17] H. H. McAdams and A. Arkin, *It's a noisy business! Genetic regulation at the nanomolar scale*, Trends Genet. 15 (1999), 65.

[18] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii and U. Alon, *Network motifs: simple building blocks of complex networks*, Science 298 (2002), 824.

[19] E. M. Ozbudak, M. Thattai, H. N. Lim, B. I. Shraiman and A. van Oudenaarden, *Multistability in the lactose utilization network of Escherichia coli*, Nature 427 (2004), 737.

[20] J. Paulsson and M. Ehrenberg, *Random signal fluctuations can reduce random fluctuations in regulated components of chemical regulatory networks*, Phys. Rev. Lett. 84 (2000), 5447.

[21] J. Paulsson, *Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation*, Proc. Natl. Acad. Sci. USA 97 (2000), 7148.

[22] J. Paulsson, *Multileveled selection on plasmid replication*, Genetics 161 (2002), 1373.

[23] J. Paulsson, *Summing up the noise in gene networks*, Nature 427 (2004), 415.

[24] M. Ptashne, *A Genetic Switch: Phage λ and Higher Organisms*, 2nd edition, Cell Press and Blackwell Scientific Publications, Cambridge, MA, 1992.

[25] N. Rosenfeld, M. B. Elowitz and U. Alon, *Negative autoregulation speeds the response times of transcription networks*, J. Mol. Biol. 323 (2002), 785.

[26] P. S. Swain, M. B. Elowitz and E. D. Siggia, *Intrinsic and extrinsic contributions to stochasticity in gene expression*, Proc. Natl. Acad. Sci. USA 99 (2002), 12795.

[27] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, North-Holland, 1992.

[28] A. M. Walczak, M. Sasai, and P. Wolynes, *Self consistent proteomic field theory of stochastic gene switches*, Biophysical Journal 88, (2005), 828.

[29] P. B. Warren and P. R. ten Wolde, *Enhancement of the stability of genetic switches by overlapping upstream regulatory domains*, Phys. Rev. Lett. 92 (2004), 128101.

[30] P. B. Warren and P. R. ten Wolde, *Chemical models of genetic toggle switches*, J. Phys. Chem. B 109 (2005), 6812.

STOCHASTIC MODELS IN BIOLOGICAL SCIENCES
BANACH CENTER PUBLICATIONS, VOLUME 80
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WARSZAWA 2008

FROM CONVERGENCE OF OPERATOR SEMIGROUPS TO GENE EXPRESSION, AND BACK AGAIN

ADAM BOBROWSKI

Institute of Mathematics, Polish Academy of Sciences
Katowice branch, Bankowa 14, 40-007 Katowice, Poland
on leave from
Department of Mathematics
Faculty of Electrical Engineering and Computer Science
Lublin University of Technology
Nadbystrzycka 38A, 20-618 Lublin, Poland
E-mail: a.bobrowski@pollub.pl

To the memory of Professor Andrzej Lasota

Abstract. The subject of the paper is reciprocal influence of pure mathematics and applied sciences. We illustrate the idea by giving a review of mathematical results obtained recently, related to the model of stochastic gene expression due to Lipniacki et al. [38]. In this model, featuring mRNA and protein levels, and gene activity, the stochastic part of processes involved in gene expression is distinguished from the part that seems to be mostly deterministic, and the dynamics is expressed by means of a piece-wise deterministic Markov process. Mathematical results pertain to asymptotic behavior of the process in time as well as limit behavior when certain parameters may be assumed to be large. These results are but an inspiration to considering the ways applied sciences influence pure mathematics by supplying fresh ideas and providing

2000 *Mathematics Subject Classification*: 47D07, 60J25, 60J35, 60J55.

Key words and phrases: contraction semigroup, generator, convex combination, approximation formulae, Trotter-Kato theorem, degenerate convergence, Feller process, weak convergence of processes, time-changed process, Feynman-Kac formula, Volkonskii's formula, gene expression, Fokker-Planck equation, asymptotic stability.

I would like to thank all those who in various ways contributed to this research. These include (in random order) T. Lipniacki, M. Kimmel, A. B. Brasier, A. Marciniak-Czochra, P. Paszek, J. Kisyński, J. Zabczyk, H. Leszczyński, R. Rudnicki, K. Pichór, K. Bogdan, T. Komorowski and W. Rzymowski. This work was partly supported by the Polish Government, grant number 4T 07A 001 30, and by TODEQ.

The paper is in final form and no version of it will be published elsewhere.

 © Instytut Matematyczny PAN, 2008

new challenges. On the other hand, they may also be seen as an exemplification of the fact that statements that seem to be almost obvious and are often taken for granted in applied sciences may require mathematical scrutiny and non-standard proofs.

1. Introduction. As the title is meant to suggest [51], this paper may be seen as a report from scientific adventure voyage I undertook around 1997 to come back home some 10 years later. I did not have any dragons to fight against, as was the case with the poor hobbit, but my story seems to be of some interest as it shows how close the two apparently distant words – these of pure, abstract mathematics and of applied sciences – really are. And how mightily they may influence each other, if only we allow a two-way traffic.

I did not make the trip by myself, and am not a pioneer in any way – I was helped by fearless guides living in the high mountains and dreadful deserts of the Lands Between, and it is to them I owe thanks for the safe passage there and back again. But the way is open, and many may go, as I did, to render help to inhabitants of the far lands and bring back the treasures of the old Smaug.

2. Home at Bag End: mathematical background

2.1. *Convergence of Laplace transforms, and the Trotter–Kato–Neveu theorem.* Let $\mathbb{X}$ be a Banach space, $f_n : [0,\infty) \to \mathbb{X}, n \geq 1$ be equibounded measurable functions in that there exists an $M \geq 0$ such that $\|f_n(t)\| \leq M$ for all n and almost all $t \in [0,\infty)$, and let $r_n : (0,\infty) \to \mathbb{X}, n \geq 1$, be the Laplace transforms of f_n :

$$r_n(\lambda) = \int_0^\infty \mathrm{e}^{-\lambda t} f_n(t)\,\mathrm{d}t.$$

By the Lebesgue Dominated Convergence Theorem, convergence of f_n implies convergence of r_n. However, as an example of $f_n(t) = \mathrm{e}^{\imath n t}$ (where $\imath = \sqrt{-1}$) shows, the converse statement is in general not true, and some additional assumptions have to be made on f_n to guarantee its validity. A common assumption of such a type is that of equicontinuity of f_n at each point $t \in [0,\infty)$. (An elegant proof of sufficiency of equicontinuity, due to Kisyński [3, 7], is based on the fact that in such a set up, $[0,\infty) \ni t \mapsto F(t) = (f_n(t))_{n\geq 1}$ is a continuous function in the space of bounded $\mathbb{X}$-valued sequences (equipped with the supremum norm). Since the Laplace transform of F (exists and) has values in the subspace of convergent sequences, subspace separation theorems imply that the values of F must be in the same subspace, giving the desired result.) Additionally, under such assumption, the limit function is continuous, and, hence, convergence is uniform on compact subintervals of $[0,\infty)$ – a mode of convergence known as *almost uniform*. Moreover, if we restrict our attention to almost uniform convergence, equicontinuity turns out to be necessary: continuous f_n converge almost uniformly on compact subintervals iff they are equicontinuous and their Laplace transforms converge [3] Thm 1.7.5. – but, certainly, it is the previous result that is of more practical value.

In this context, the Trotter–Kato–Neveu theorem, one of the corner stones of the theory of one-parameter operator semigroups, is at least surprising. It says that a sequence of equibounded C_0 semigroups $\{T_n(t), t \geq 0\}$, $n \geq 1$ converges strongly (*i.e.* in the

sense of the norm in $\mathbb{X}$) to a limit C_0 semigroup $\{T(t), t \geq 0\}$ iff the Laplace transforms $R_{\lambda,n} = \int_0^\infty \mathrm{e}^{-\lambda t} T_n(t)\,\mathrm{d}t$ converge to $R_\lambda = \int_0^\infty \mathrm{e}^{-\lambda t} T(t)\,\mathrm{d}t$, and convergence of semigroups is almost uniform in $[0, \infty)$ – apparently, no assumption of equicontinuity is needed here. (Equiboundedness means that there exists a constant $M > 0$ such that $\|T_n(t)\| \leq M$ for all $t \geq 0$ and $n \geq 1$.) But of course, as we shall see soon, this is one of the cases of "hidden equicontinuity", where equicontinuity is implicit in other assumptions, like equiboundedness of derivatives.

Let us recall that a C_0 semigroup (or: a strongly continuous semigroup) of operators is a family of bounded operators $\{T(t), t \geq 0\}$ in $\mathbb{X}$ such that:

- $T(t)T(s) = T(t+s), s, t \geq 0$,
- $T(0) = I_{\mathbb{X}}$,
- $\lim_{t \to 0+} T(t) = I_{\mathbb{X}}$ (strongly).

As developed in the well-known and beautiful theory (see e.g. [3, 20, 22, 27, 44, 52] etc.) such families have very nice properties that make them of importance and great applicability [3, 8, 19, 20, 44] etc; we name three of them. The first is that for all $x \in \mathbb{X}$, the map $[0, \infty) \ni t \mapsto T(t)x$, called the trajectory of the semigroup, is strongly continuous. The second is that the set $D(A)$ of x for which the right-hand derivative at $t = 0$ of related trajectory exists, is dense in $\mathbb{X}$. Moreover, the operator of this derivative,

$$Ax = \lim_{t \to 0} \frac{1}{t}(T(t)x - x)$$

defined on $D(A)$ is closed, and for $x \in D(A)$, the related trajectories are differentiable with

$$\frac{\mathrm{d}T(t)x}{\mathrm{d}t} = AT(t)x = T(t)Ax. \tag{1}$$

A is called the generator of the semigroup. In other words, trajectories of the semigroup starting at elements of $D(A)$ are solutions to the Cauchy problem for its generator. Conversely, if a Cauchy problem for a densely defined operator A is well-posed in that solutions exist for initial conditions $x \in D(A)$ and depend continuously on the initial conditions, then A is the generator of a semigroup. (The Cauchy problem for a, say, linear operator A and initial condition x_0 is that of existence and uniqueness of solutions to equation $x(t) = Ax(t), t \geq 0$ with $x(0) = x_0$.)

The third property is that for any semigroup there exist constants $M \geq 1$ and $\omega \in \mathbb{R}$ such that $\|T(t)\| \leq M\mathrm{e}^{\omega t}$, and, consequently, the Laplace transform $R_\lambda x = \int_0^\infty \mathrm{e}^{-\lambda t} T(t)x\,\mathrm{d}t$ exists for all $\lambda > \omega$. It turns out that all $\lambda > \omega$ belong to the resolvent set of A and that $R_\lambda = (\lambda - A)^{-1}$.

It is these properties that force equicontinuity in the set up of the Trotter–Kato–Neveu theorem. To see this, let us be more precise as to what is assumed there. We assume namely that $R_{\lambda,n}$ converge and that the limit operators R_λ are of quite a special form: they are to be equal to the resolvent of the generator of the limit semigroup, and hence, in particular they must have a common range, equal to $D(A)$, which is dense in $\mathbb{X}$. How this comes into play is seen from the following proof of the Trotter–Kato–Neveu theorem, which goes back to T.G. Kurtz [35]. Let us take $x \in D(A)$ and $\lambda > 0$. Then, there exists a $y \in \mathbb{X}$ such that $x = (\lambda - A)^{-1}y$. Hence, $x_n = R_{\lambda,n}y = (\lambda - A_n)^{-1}y$ belongs to the domain

$D(A_n)$ of the infinitesimal generator A_n of the $n-$th semigroup, and $\lim_{n\to\infty} x_n = x$, by assumption. On the other hand, $A_n x_n = \lambda(\lambda - A_n)^{-1} y - y = \lambda x_n - y$ converges to $\lambda x - y$. Therefore, $\sup_n \|A_n x_n\| < \infty$. Consequently, by (1), the derivatives of $[0,\infty) \ni t \mapsto T_n(t)x_n$ are equibounded, and so these maps are equicontinuous, even more, they are Lipschitz continuous with the same constant. Since their Laplace transforms converge to the Laplace transform of $[0,\infty) \ni t \mapsto T(t)x$, the functions must converge to this function (almost uniformly). Since x_n converges to x and the semigroups are equibounded, the same is true of $[0,\infty) \ni t \mapsto T_n(t)x$. Using equicontinuity of the semigroups, $D(A)$ being dense in $\mathbb{X}$, we obtain convergence for all $x \in \mathbb{X}$.

2.2. *Trouble with generalizations of the Trotter–Kato–Neveu theorem: degenerate convergence.* A serious drawback of the Trotter–Kato–Neveu theorem is that it assumes a priori existence of the limit semigroup $\{T(t), t \geq 0\}$. In a more natural set up the limit semigroup is not given and we know merely that $R_{\lambda,n}$ converge:

$$\lim_{n\to\infty} R_{\lambda,n} =: R_\lambda, \qquad \lambda > 0.$$

Since the $R_{\lambda,n}$ are resolvents of generators A_n of $\{T_n(t), t \geq 0\}$, they satisfy the so-called Hilbert equation: $(\lambda - \mu)R_{\lambda,n}R_{\mu,n} = R_{\mu,n} - R_{\lambda,n}$ for $\lambda, \mu > 0$, and so must R_λ :

$$(\lambda - \mu)R_\lambda R_\mu = R_\mu - R_\lambda, \quad \lambda, \mu > 0.$$

This is to say that $R_\lambda, \lambda > 0$ is a pseudoresolvent; in particular, the operators R_λ have common range and kernel.

In general, as it will be clear from examples to be presented below, R_λ is not a resolvent of a single operator. However, if the range of R_λ is dense in $\mathbb{X}$, we may construct a semigroup $\{T(t), t \geq 0\}$ with generator A satisfying $\int_0^\infty \mathrm{e}^{-\lambda t} T(t)\,\mathrm{d}t = (\lambda - A)^{-1} = R_\lambda$. This semigroup then turns out to be the limit semigroup of $\{T_n(t), t \geq 0\}$. If the range of R_λ is not dense in $\mathbb{X}$, the semigroup as above may be constructed merely on

$$\mathbb{X}_0 = cl(Range R_\lambda).$$

More specifically, in $\mathbb{X}_0$ there exists an operator A such that all $\lambda > 0$ belong to its resolvent set and $(\lambda - A)^{-1}x = R_\lambda x$, for $x \in \mathbb{X}_0$. This operator may be shown to generate a strongly continuous semigroup $\{T(t), t \geq 0\}$ such that

$$\lim_{n\to\infty} T_n(t)x = T(t)x, \qquad x \in \mathbb{X}_0, t \geq 0,$$

almost uniformly in $[0,\infty)$. It is worth stressing that $\mathbb{X}_0$ is *the* subspace where the semigroups $\{T_n(t), t \geq 0\}$ converge almost uniformly in $[0,\infty)$. In other words, outside of $\mathbb{X}_0$ the semigroups either do not converge at all, or do converge but not almost uniformly in $[0,\infty)$.

By the way, let us note that the proof of the Trotter–Kato–Neveu theorem presented in the previous section was based on the fact that we were able to find a large class of $x \in \mathbb{X}$ such that there existed $x_n \in D(A_n)$ with $\lim_{n\to\infty} x_n = x$ and $A_n x_n$ convergent. It turns out that all information concerning almost uniform convergence of semigroups $\{T_n(t), t \geq 0\}$ is hidden in the relation $A_{ex} \subset \mathbb{X} \times \mathbb{X}$ defined as $(x, y) \in A_{ex}$ iff there exists $x_n \in D(A_n)$ with $\lim_{n\to\infty} x_n = x$ and $\lim_{n\to\infty} A_n x_n = y$, often termed *extended limit of* A_n. (In general this is not an operator, since given x there may be many y such that

$(x, y) \in A_{ex}$.) A close inspection of Kurtz's proof of the Trotter–Kato–Neveu theorem (see the previous subsection) shows two facts. First of all, $D(A_{ex})$, defined as the set of $x \in \mathbb{X}$ such that $(x, y) \in A_{ex}$ for some y, is contained in $\mathbb{X}_0$. Secondly, if A is the generator of the limit semigroup defined in $\mathbb{X}_0$, then $D(A) \subset D(A_{ex})$. Therefore, $\mathbb{X}_0$ is the closure of $D(A_{ex})$. Moreover, it may be shown that the resolvents $R_{\lambda,n}$ converge iff for some $\lambda > 0$ the set of elements of the form $\lambda x - y$ where $(x, y) \in A_{ex}$ is dense in $\mathbb{X}$. Finally, the part of A_{ex} in $\mathbb{X}_0$, i.e. the relation $A_{ex} \cap (\mathbb{X}_0 \times \mathbb{X}_0)$ is the graph of an operator in $\mathbb{X}_0$ – this operator is the generator of the limit semigroup $\{T(t), t \geq 0\}$ in $\mathbb{X}_0$. All the above facts constitute the Sova-Kurtz version of the Trotter–Kato–Neveu theorem – see [8, 19, 35, 49].

Coming back to our main thought: outside of $\mathbb{X}_0$ the semigroups may not converge, and there are plenty of examples of such behavior. For instance, the semigroups in $\mathbb{X} = \mathbb{C}$, given by $T_n(t)x = \mathrm{e}^{\imath tn}x$ we have encountered at the beginning of this paper. Also, the translation semigroups $T_n(t)x(\tau) = x(\tau - nt)$ for $\tau \geq nt$ and $= 0$ for $\tau < nt$, in the space of continuous functions with $x(0) = \lim_{\tau\to\infty} x(\tau) = 0$. In both cases, $\mathbb{X}_0 = 0$ and outside of this space the semigroups do not converge (see [6] or [8] for details). Still another example is given by the Yosida approximation of the operator $x \mapsto -x'$ (the derivative) defined in the space of continuous functions on $[0, \infty]$, with domain composed of x from this space such that $x(0) = 0$ and x' belongs to this space – see [8] Example 8.2.10.

More interestingly, the semigroups $\{T_n(t), t \geq 0\}$ *may* converge outside of $\mathbb{X}_0$, but certainly not almost uniformly in $[0, \infty)$. And in fact, as we will yet see, quite often they do. (To be sure, many examples of convergence of semigroups known in the literature, both in pure and applied mathematics, are instances of such convergence, this is particularly so with the singular perturbations and parabolic approximations to hyperbolic problems – see e.g. [2, 4, 5, 28, 40].) However, there are few general theorems that can be used to prove such convergence in practice.

To see where the difficulty lies, note that if the semigroups converge in this way, in the limit we still obtain a semigroup, an extension of the semigroup $\{T(t), t \geq 0\}$ acting in $\mathbb{X}_0$, but one that is not of class C_0. In fact, denoting it still by $\{T(t), t \geq 0\}$, by definition of $\mathbb{X}_0$, we have $\lim_{t\to 0} T(t)x = x$ merely for $x \in \mathbb{X}_0$, and so the third point of the definition of C_0 semigroup fails. As a consequence, the limit semigroup does not have nice properties enjoyed by C_0 semigroups. Yes, that is true that being measurable (as a limit of measurable semigroups) it must be strongly continuous for $t > 0$, by [27] Thm 3.5.4, but still is not of class C_0. In particular, the set of x for which the related trajectory is differentiable is not dense in $\mathbb{X}$, we cannot do the trick with equiboundedness of derivatives leading to equicontinuity, and need to look for more subtle methods.

To be sure, convergence for $x \notin \mathbb{X}_0$ *cannot* – by nature – be proved using the Trotter–Kato–Neveu theorem, and $\mathbb{X}_0$ may be pretty small or insignificant – see [11].

2.3. *Examples.* Let us turn to examples. For the first of these, given a semigroup $\{S(t), t \geq 0\}$ with $\|S(t)\| \leq M, t \geq 0$ for some $M > 0$, we define $T_n(t) = \mathrm{e}^{-nt}S(t)$. Then, $\{T_n(t), t \geq 0\}, n \geq 1$ are equibounded and certainly $\lim_{n\to\infty} T_n(t) = 0, t > 0$. On the other hand, we have $R_{\lambda,n} = \int_0^\infty \mathrm{e}^{-(\lambda+n)t}S(t)\,\mathrm{d}t \to 0$, as $n \to \infty$. Hence, $\mathbb{X}_0 = \{0\}$

and all the Trotter–Kato–Neveu theorem shows is that $0 = \mathrm{e}^{-nt}S(t)0$ converges to 0. Without realization that this theorem deals merely with convergence that is almost uniform in $[0,\infty)$ it could be found really amazing that convergence that is completely obvious by inspection cannot be captured by means of it.

For a next example, consider the most general time-continuous two-state Markov chain. Its intensity matrix is given by $\begin{pmatrix} -a & a \\ b & -b \end{pmatrix}$ where $a, b \geq 0$ are given constants. Recall that such a chain starting from state 1, waits for an exponential time of parameter a to jump to state 2. While at 2, it waits for an exponential time of parameter b, before jumping to state 1. Matrix $A_n = \begin{pmatrix} -a & a \\ b & -b-n \end{pmatrix}$ describes a modified chain, which at 1 behaves as before, but at state 2 it waits for exponential time with parameter $b+n$ and jumps to state 1 with probability $\frac{b}{b+n}$ or disappears form the state-space with probability $\frac{n}{b+n}$. Hence, putting $T_n(t) = \exp A_n t$, we should have

$$\lim_{n\to\infty} T_n(t)\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \mathrm{e}^{-at}x \\ 0 \end{pmatrix}, \qquad x, y \in \mathbb{R}, t > 0. \tag{2}$$

($T_n(t)$ are contractions in $\mathbb{R}^2$ equipped with the norm $\|(x,y)\| = |x| + |y|$.) To prove this, calculating $R_{\lambda,n} = \frac{1}{\lambda(a+b)+n(\lambda+a)}\begin{pmatrix} b+n & a \\ b-\lambda & \lambda+a \end{pmatrix}$ we see that it converges to $\begin{pmatrix} \frac{1}{\lambda+a} & 0 \\ 0 & 0 \end{pmatrix}$. Hence, $\mathbb{X}_0 = \mathbb{R}\times\{0\}$ and the Trotter–Kato–Neveu Theorem gives (2) for $y = 0$. In particular, it does not and may not provide any information for $y \neq 0$. Fortunately, the full result may be obtained from the explicit form of $T_n(t)$ given e.g. in [42], see [11] for details.

For our third example, suppose we are given $\epsilon > 0$ and the generator A of a bounded (and strongly continuous) cosine operator function. (Cosine operator functions are related to the second order Cauchy problem $x''(t) = Ax(t), x(0) = y, x'(0) = 0$ in the same way as the semigroups are related to the first order Cauchy problem. An operator A generates a strongly continuous cosine operator function iff the related second order Cauchy problem is well-posed.) Consider the abstract telegraph equation with small parameter:

$$\epsilon x_\epsilon''(t) + x_\epsilon'(t) = Ax_\epsilon(t), \qquad x_\epsilon(0) = y, x_\epsilon'(0) = z. \tag{3}$$

Formally, letting $\epsilon \to 0$, we obtain the abstract heat equation:

$$x_0'(t) = Ax_0(t), \qquad x_0(0) = y,$$

and so, our intuition is that $\lim_{\epsilon\to 0} x_\epsilon(t) = x_0(t)$. When we turn to derivatives, we may still expect that $\lim_{\epsilon\to 0} x_\epsilon'(t) = x_0'(t)$ for $t > 0$, but we cannot obtain convergence at $t = 0$ unless $z = Ay$, for we have $x_\epsilon'(0) = z$ and $x_0'(0) = Ay$.

Using the main theorem of [34], telegraph equation (3) may be transformed to an equivalent first-order system of equations, and then the claim concerning convergence of solutions and derivatives with initial conditions $x_\epsilon(0) = y, x_\epsilon'(0) = Ay$ follows from the Trotter–Kato–Neveu theorem – see [6]. The proof for $x_\epsilon'(0) \neq Ay$ requires more subtle methods. In [6] we use an explicit representation of solutions of (3) in terms of related stochastic processes, and the continuity theorem. In [11] we show the desired result by comparing the related semigroups with semigroups that are easier to show to be

convergent. Still another method is presented in [33]. (Of course, diffusion approximation to telegraph equation has been studied in many contexts by a number of authors – see *e.g.* the bibliography in [5], [6] and [28]. Here, we merely stress its semigroup-theoretic implications.)

Our final example, also taken from [6], concerns convergence of Brownian motions in $[0,\infty)$ with changing behavior at the screen $\tau = 0$. Namely, we consider $\mathbb{X} = C[0,\infty]$ and the operators $A_\epsilon, \epsilon > 0$ defined as $A_\epsilon = \frac{\mathrm{d}^2}{\mathrm{d}\tau^2}$ with $\mathcal{D}(A_\epsilon)$ composed of $x \in \mathbb{X}$ such that $\frac{\mathrm{d}^2 x}{\mathrm{d}\tau^2}$ exists and belongs to $\mathbb{X}$, and $x(0) = \epsilon x'(0)$. These operators are generators of Feller semigroups $\{T_\epsilon(t), t \geq 0\}$ in $C[0,\infty]$ related to Brownian motions that upon touching the boundary where $\tau = 0$ are partly reflected and partly disappear from the space – see [29, 41, 46]. The $\epsilon > 0$ is a coefficient of elasticity; it decides what happens to the Brownian particle touching the screen. The smaller the ϵ the larger is the fraction of particles that are being annihilated at 0; the remaining ones are reflected at the screen and continue their motion. From this description, and from the form of the generators, it is intuitively clear that as $\epsilon \to 0$, these processes approach the minimal Brownian motion, where at the screen $\tau = 0$ the particle is instantly annihilated and disappears from the space; the semigroup $\{T_0(t), t \geq 0\}$ related to the minimal Brownian motion is defined in the subspace $\mathbb{X}_0$ of $\mathbb{X}$ where $x(0) = 0$, and its infinitesimal generator is $A_0 x = \frac{1}{2}x''$, on the natural (maximal) domain.

The fact that $\lim_{\epsilon\to 0} T_\epsilon(t)f = T_0(t)f$ for $f \in \mathbb{X}_0$ may be proved using the Trotter–Kato–Neveu theorem. This theorem, however, does not give information on $\lim_{\epsilon\to 0} T_\epsilon(t)f$ for $f \notin \mathbb{X}_0$. To treat the latter case, we consider $C[0,\infty]$ as the space of complex functions and note that the semigroups $\{T_\epsilon(t), t \geq 0\}$ are uniformly analytic, and express them with the help of the Dunford integral – see [6] for details.

3. From pure to applied mathematics: gene expression. In 1994, when I defended my PhD thesis with Prof. J. Kisyński as the thesis advisor, examples such as presented above called for a general treatment, but none seemed to be available. For some time I enjoyed finding connections of these problems with the theory of non-densely defined operators, in particular integrated semigroups, but gradually got persuaded I do not have much fresh inspiration to deal with them. The subject seemed to be more and more dry and uninteresting, and I decided I need to start doing things that are related to "real life" and lead to "real applications", little though I new about them. I became interested in biological models.

As a result of my pursuit and a punishment for my folly, I was thrown into the dungeon (actually: the second floor) of a building where there was no mathematician to be found: just geneticists. For a number of years I become involved with studying mathematical models of one of the main population genetics forces, called genetic drift, and in particular became familiar with the beautiful process of Kingman's coalescence – see [32]. Of course, being a mathematician, I tended to use mathematical language I was familiar with to describe the objects from biology, and used mathematical tools to deal with the problems coming from this field – to be sure, it is fun. I never thought, though, I would ever come back to "degenerate convergence of semigroups."

But then, in May 2005 I was asked for my opinion on the then to-be-published model of stochastic gene expression due to Lipniacki et al. [38], see also [39] and [25].

To explain the model, let us recall that gene expression is the process in which gene's DNA sequence is converted into the structures and functions of a cell (most often: proteins or RNA). The idea behind this notion is apparently that the gene itself is invisible, or at least *was* invisible at the time when the term was coined, but expresses its existence via its visible products, such as proteins. The process itself is quite complicated and there are significant differences in gene expression between eukaryote and bacteria. Even if we restrict ourselves to eukaryote, the process will still depended on a kind of cell, phase of its development, and on metabolical and physiological state.

Fortunately, there are some stages of gene expression which are more or less common to all eukaryote, which in an extremely simplified way may be described as follows. First, we must recall that (almost) all genetic information is contained in the nucleus of a cell where cell's DNA is, and the process of gene expression does not start until for this or that reason the gene becomes active. At this point, information contained in DNA is transcribed into so-called messenger RNA (mRNA), a medium of transportation. Then, mRNA molecules cary the information to the cytoplasm, where it is translated into protein sequence. After translation, the resulting protein may still be modified and then travels to its destination (target) in the cell. The last stages are natural degradation of mRNA and degradation of mRNA.

It should be noted here that typically there are several copies of a single gene in the genome. (Eukaryotic cells are diploidal, i.e. typically have two copies of each gene in their genome.) It seems reasonable to assume that they become active and inactive independently of each other. However, it is biologically established that the level of proteins influences intensities of activation and inactivation of the copies.

Authors of [38] argue that while the process of transcription, transportation and translation are more or less deterministic, the process of gene activation (inactivation) is of stochastic nature, basically due to a limited number reactants. This, by the way, is one of the main subjects and important input of the paper. Moreover, they propose a mathematical model where the state of the process is described by three variables (x, y, γ) where x is the mRNA level, y is the protein level, and γ is the number of active gene copies. After some simplifying and re-scaling procedure, the dynamics is expressed in the form of the following system of stochastic ordinary differential equations:

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= \gamma - x, \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= r(x - y), \end{aligned} \tag{4}$$

where $r > 0$ is a given constant, and γ is an integer-valued random process with intensity of increase by 1 equal to $a\alpha(x, y)$ and intensity of decrease by 1 equal to $b\beta(x, y)$ for some non-negative functions α and β and positive constants a and b. (Biologically, it is clear that α and β do depend on y, but not on x. Introducing assumption on dependence on x to the model does not, however, lead to additional mathematical difficulties.) Such a model captures all the important features of gene expression. More specifically, the first

equation shows that when all gene copies are inactive, *i.e.* when $\gamma = 0$, the number of mRNA decreases exponentially, and that an increase in γ results in the gradual growth of x. On the other hand, by the second equation, a burst (or a decrease in number) of mRNA molecules produces a burst (or a decrease in number) of protein molecules, with protein level y "catching up" with mRNA level with efficiency r. Finally, the level of mRNA and protein influences probability of the change in the number of active gene copies.

I need to confess, the example offered by this model is a very bad one, pedagogically. For years I was told, and got used to the fact that in applied mathematics, for each problem you face, an appropriate mathematical language with appropriate tools must be chosen. That, in other words, you should not stick to your favorite theory, but always learn new ones to match the problem you are to solve. But for me, mathematically, this model was exactly what I specialized in and what I was interested in for years. Of course, I had the advantage of working on a model that was already prepared by somebody else. But still, it seemed to suit my interests too well.

Actually the number of mathematical connections of the model I became aware of is quite surprising. First of all, it is clear that where it not for the fact that the intensities of jumps of γ depend on y (and possibly on x), the process (x, y, γ) would have been an example of *random evolution*, i.e. a member of a class of stochastic processes introduced by Griego and Hersh [23, 24] after Kac's famous work on stochastic solutions of the telegraph equation [30]. After I consulted with prof. J. Zabczyk, the process turned out to be an example of a piece-wise deterministic process of M.H.A. Davis [14, 15, 16], who prophesied in 1984: "These (processes) were only isolated rather recently but seen general enough to include as special cases practically all the non-diffusion continuous-time processes of applied probability" ([15], p. i).

Secondly, the authors of [38] were interested in the the Fokker-Planck-type equation for the process $(x(t), y(t), \gamma(t)), t \geq 0$, which we write here in the case of a single gene:

$$\begin{aligned} \frac{\partial f_0}{\partial t} + \frac{\partial}{\partial x}(-x f_0) + r\frac{\partial}{\partial y}((x-y)f_0) &= b\beta f_1 - a\alpha f_0, \\ \frac{\partial f_1}{\partial t} + \frac{\partial}{\partial x}((1-x) f_1) + r\frac{\partial}{\partial y}((x-y)f_1) &= a\alpha f_0 - b\beta f_1; \end{aligned} \tag{5}$$

here $f_i(t, \cdot, \cdot), i = 0, 1$ are parts of a density of the process at time $t \geq 0$ related to $\gamma = 0, 1$, respectively. (Let us note in passing that were the intensities α and β equal to zero, the above system would have been uncoupled and would describe two independent motions along trajectories of solutions of (4) with fixed γ equal either 0 or 1, by means of a family of Frobenius–Perron operators in a related L^1 space – see [36]; the non-zero right-hand side describes the rules of jumps between states $\gamma = 0$ and $\gamma = 1$.) In particular one of the questions was that of existence (and form) of densities that would be invariant for the equation. Certainly, being A. Lasota's student, I was familiar with the theory of Markov semigroups in L^1 that provides the means for approaching such questions – roughly speaking, Markov semigroups are all about dynamics of densities of Markov processes, being composed of operators that to an initial density of a Markov process assign its densities at times $t > 0$; see [36]. As a result of another student of Lasota's,

namely R. Rudnicki's expertise in this field, existence of such a density and convergence of all other densities to this one as $t \to \infty$ was proved in [12].

But these were two hypotheses concerning limit behavior of (4) as a, b and r tend to infinity, that really captured my attention, as they proved to be hypotheses on degenerate convergence of semigroups. Parenthetically, let us recall what Einar Hille wrote in 1948 [26]: "I hail a semigroup when I see one and I seem to see them everywhere. (Friends have observed, however, that there are mathematical objects which are not semigroups.)" In a sense, thus, it is not surprising that I also see them even in a model of gene expression. Nevertheless, I am still amazed that questions from abstract mathematics may be of relevance in applied sciences.

The first hypothesis (not discussed in the published version of Lipniacki's paper) concerned the limit behavior of (4) when $r \to \infty$. Heuristically, as $r \to \infty$, the efficiency with which y chases x, increases infinitely, and so in the limit, we should expect that y catches up with x immediately after any change. In other words, we should have $x = y$ all the time and (4) should reduce to a single stochastic equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \gamma - x, \tag{6}$$

where γ is a random process with values in $\{0, 1\}$, with activation intensity $a\tilde{\alpha}(x) = a\alpha(x, x)$ and inactivation intensity $b\tilde{\beta}(x) = b\beta(x, x)$. Biologically, r is defined as the ratio of protein and mRNA degradation rates. Hence, large r corresponds to the case where mRNA is much more stable than the corresponding protein. In such a scenario, the mRNA's level dictates the protein's level: compare [45] Fig. 7A (R=0.78) and Fig. 7D (R=0.92). This is the case with protein IκBα in the NFκB module (in the absence of A20 protein) – see [37], and with protein p53 – see [13].

The second hypothesis concerned the limit behavior of (4) as a and b tend to infinity in such way that a/b tends to a positive constant c. Lipniacki argues that in such a case the jumps occur infinitely often and γ reaches statistical equilibrium. Hence, in (4) it may be replaced by its conditional expected value (given x and y). In other words, in the limit, (4) becomes

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= \frac{c\alpha}{c\alpha + \beta} - x, \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= r(x - y). \end{aligned} \tag{7}$$

To explain it in more detail: the paper [38] was concerned mainly with the situation where transitions between active and inactive states of the gene are relatively rare, so that to explain, in particular, bursts in production of mRNA reported in [21]. Passage to the limit as a and b tend to infinity corresponds to the case of frequent transitions studied already in the 1980' – see [1, 48], the case that can be called classical today. See [50] where frequent transitions are assumed tacitly, especially Fig. 3 where the transition rate is very similar to the term $\frac{c\alpha}{c\alpha+\beta}$ appearing in (7); for an analogous formulae for transition rate in a slightly different scenario see [13]. Similar passages to the limit may be found in [31]; in particular, passage from (B1) to (B20), from (34) to (39), and from (C1)-(C3) to (46)-(47). See also [43].

When put in the context of weak convergence of related stochastic processes, these hypotheses are statements on convergence of certain conditional expected values, as explained below. First, let us note that in the case of a single gene copy, *i.e.* the case where γ belongs to $\{0,1\}$, it suffices to consider (4) for x and y in $[0,1]$. Indeed, all solutions to these equations starting from outside of the unit square $\mathcal{Q}=[0,1]^2$ tend to this square, and upon reaching it stay there for ever. Hence, we may consider $(x(t),y(t),\gamma(t)), t\geq 0$ as a process with values in two copies of $\mathcal{Q}$ (one corresponding to $\gamma=0$, the other to $\gamma=1$). A continuous function on the state space may therefore be identified with two functions, say f_0 and g_0, on $\mathcal{S}$. For $t\geq 0$, we consider

$$\begin{aligned} f(t,x,y) &= \mathbb{E}_{x,y,0}\left\{f_0(x(t),y(t))1_{\{\gamma(t)=0\}} + g_0(x(t),y(t))1_{\{\gamma(t)=1\}}\right\} \\ g(t,x,y) &= \mathbb{E}_{x,y,1}\left\{f_0(x(t),y(t))1_{\{\gamma(t)=0\}} + g_0(x(t),y(t))1_{\{\gamma(t)=1\}}\right\}, \end{aligned} \tag{8}$$

where $\mathbb{E}_{x,y,i}, i=0,1$ is the conditional expectation, given that $x(0)=x, y(0)=y$ and $\gamma(0)=i$. Since the process $(x(t),y(t),\gamma(t)), t\geq 0$ may be shown to be Feller (see Section 4 for definition), these functions are continuous, and if f_0 and g_0 are sufficiently regular (for example of class C^1), they satisfy the so-called Kolmogorov equation:

$$\begin{aligned} \frac{\partial f}{\partial t} &= \quad -x\frac{\partial f}{\partial x} + r(x-y)\frac{\partial f}{\partial y} - a\alpha f + a\alpha g, \\ \frac{\partial g}{\partial t} &= (1-x)\frac{\partial g}{\partial x} + r(x-y)\frac{\partial g}{\partial y} - b\beta g + b\beta f, \end{aligned} \tag{9}$$

with initial conditions $f(0,x,y)=f_0(x,y)$ and $g(0,x,y)=g_0(x,y)$, in a sense dual to (5).

Weak convergence of processes involved in the first hypothesis is then equivalent to convergence of solutions of (9) to solutions of the Kolmogorov equation related to (6), which reads

$$\begin{aligned} \frac{\partial f}{\partial t} &= \quad -x\frac{\partial f}{\partial x} - a\tilde{\alpha} f + a\tilde{\alpha} g, \\ \frac{\partial g}{\partial t} &= (1-x)\frac{\partial g}{\partial x} - b\tilde{\beta} g + b\tilde{\beta} f. \end{aligned} \tag{10}$$

Here, $f(t,\cdot)$ and $g(t,\cdot)$ are functions of one variable $x\in[0,1]$, and it should be noted that such functions may be identified with continuous functions on $\mathcal{S}$ that do not depend on y. In other words, we have another, and quite natural, example of a family of C_0 semigroups in $C(\mathcal{Q})\times C(\mathcal{Q})$ approximating a C_0 semigroup acting only in a subspace of this space – the subspace that is isometrically isomorphic to $C[0,1]\times C[0,1]$. And again, what we may prove with the help of the Trotter–Kato theorem is that if we choose initial conditions of (9) not to depend on y, then in the limit as $r\to\infty$ we will obtain solutions to (10), with the same initial condition. This, however, does not prove weak convergence of related processes. To establish weak convergence, we need to prove a similar result for any initial conditions. In [9] we show that if we start from any $(f_0,g_0)\in C(\mathcal{Q})\times C(\mathcal{Q})$, then for all $t>0$ in the limit we obtain solutions of (10) starting from $x\mapsto f_0(x,x)$ and $x\mapsto g_0(x,x)$ – the proof is based on the Phillips perturbation theorem.

The situation is similar with the second hypothesis. In terms of weak convergence of related processes it states that as a and b tend to infinity in the prescribed way, the

solutions to (9) converge to the solutions of the equation

$$\frac{\partial f}{\partial t} = \left(\frac{c\alpha}{c\alpha+\beta} - x\right)\frac{\partial f}{\partial x} + r(x-y)\frac{\partial f}{\partial y}. \tag{11}$$

It is worth noting that the limit process described by (7) is deterministic: (7) is a system of autonomous ordinary differential equations. Solutions to this system exist under suitable, classic assumptions on α and β, and form a semi-flow, and solutions to (11) may be conveniently expressed through this semi-flow. What is important is that again, (11) is about dynamics of a single function form $C(\mathcal{Q})$, while (9) is about dynamics of pairs of functions from this space. In other words, if we identify elements of $f \in C(\mathcal{Q})$ with pairs (f, f) in $C(\mathcal{Q}) \times C(\mathcal{Q})$, we have another example of a family of C_0 semigroups in $C(\mathcal{Q}) \times C(\mathcal{Q})$ approximating a semigroup that acts merely on the subspace of $C(\mathcal{Q}) \times C(\mathcal{Q})$; this subspace is isometrically isomorphic to $C(\mathcal{Q})$. What the Trotter–Kato–Neveu theorem shows is that if we choose initial conditions of (9) in such a way that $f_0 = g_0$, then both f and g in this equation converge to the solution of (11) starting from f_0 – see [9]. But much more is true, and this cannot be proved with the help of the Trotter–Kato–Neveu theorem. Specifically, in [9] we also prove that taking any f_0 and g_0, both parts of the solution of (9) at $t > 0$ tend to the solution of (11) starting from $\frac{\beta}{c\alpha+\beta}f + \frac{c\alpha}{c\alpha+\beta}g$.

4. From applied to pure mathematics. An afterthought to the story described above is as follows. Even if we treat applied mathematics as a tool to prove hypotheses of non-mathematicians, who gains more by taking part in cooperation, a non-mathematician who's hypothesis is substantiated, or a mathematician who deals with a non-trivial problem (with a "real" application)? To be sure, I am not easily thrilled by the word "applied". I am persuaded that there are as many worthless papers published in applied sciences as there are in pure mathematics. I agree also that mathematics behind many biological problems, even those really relevant, is simply trivial – not all problems interesting for biologists or physicists are interesting for mathematicians. On the other hand, however, intuitions born outside of mathematics may become an invaluable source of inspiration to mathematicians, and an incentive to develop their theories to cover cases that are relevant for other sciences.

Just to say the least: the proof of asymptotic stability of (5) required a reconsideration of the methods available for such problems and lead to their further improvement and development – especially dependence of α and β on y (and possibly on x), natural from biological point of view, constituted a substantial difficulty – see [12]. Similarly with examples of degenerate convergence of semigroups: for several months they were a substantial challenge and lead me to a further research on limitations of the Trotter–Kato–Neveu theorem – see [11].

But I would like to conclude with another example of the way intuitions of non-mathematicians may lead to establishing new theorems of pure mathematics. This example concerns convex combinations of Feller generators, and is another by-product of Lipniacki's second hypothesis.

To explain the idea, let us recall that a Markov process $p(t), t \geq 0$ with values in a locally compact space $\mathcal{S}$ is said to be a Feller process ([8, 19, 20, 46]) if the corresponding

transition semigroup

$$T(t)f(p) = \mathbb{E}_p\, f(p(t)), \quad p \in \mathcal{S}, t \geq 0, f \in C_0(\mathcal{S}), \tag{12}$$

where $\mathbb{E}_p$ is the conditional expectation conditional on $p(0) = p$, is a strongly continuous semigroup in $C_0(\mathcal{S})$; in particular, we assume that $T(t)f$ belongs to $C_0(\mathcal{S})$ for $f \in C_0(\mathcal{S})$ and $t \geq 0$ – the reader will notice that (8) is a particular case of (12). Isolated and studied first by Feller himself, Feller semigroups (i.e. semigroups related to Feller processes) are comparatively easily treatable analytically and are best described (see references given above) – in fact, we could claim their theory is by all means classical.

It is clear that all information on finite-dimensional distributions of the process is then hidden in the related semigroup. On the other hand, such a semigroup is fully characterized by its infinitesimal generator, or by the related Cauchy problem. Now, (9) is a Cauchy problem for such a process with quite "compound" generator. In particular it involves operators, say $B_i = (i - x)\frac{\partial}{\partial x} + r(x - y)\frac{\partial}{\partial y}$ which are generators of processes of deterministic motions along integral curves of ordinary differential equations:

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= i - x, \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= r(x - y), \end{aligned} \tag{13}$$

and its remaining part is responsible for jumps between two copies of $\mathcal{Q}$. It is interesting to note that equation (11) may be obtained from (9) by letting $f = g$, and adding the first equation multiplied by $\frac{\beta}{c\alpha+\beta}$ to the second equation multiplied by $\frac{c\alpha}{c\alpha+\beta}$. Moreover, $\frac{\beta}{c\alpha+\beta}$ and $\frac{c\alpha}{c\alpha+\beta}$ are clearly the probabilities that the gene activity process γ being at statistical equilibrium, equals 0 and 1, respectively. In other words the generator of the limit, deterministic process of movement along integral curves of (7) is a convex combination of operators B_i, and that process may be seen as a result of averaging motions along integral curves of (13), with weights $\frac{\beta}{c\alpha+\beta}$ and $\frac{c\alpha}{c\alpha+\beta}$.

This suggests the following, more abstract result: let, as above, $\mathcal{S}$ be a locally compact Hausdorff space, and let α and β be two non-negative continuous functions on $\mathcal{S}$ with $\alpha+\beta = 1$. Furthermore, let A, B and C be generators of Feller semigroups $\{T_A(t), t \geq 0\}$, $\{T_B(t), t \geq 0\}$, and $\{T_C(t), t \geq 0\}$ in $C_0(\mathcal{S})$ with related Feller processes $\{X_A(t), t \geq 0\}$, $\{X_B(t), t \geq 0\}$ and $\{X_C(t), t \geq 0\}$. Finally, suppose that $\mathcal{D}(A) \cap \mathcal{D}(B)$ is a core for C and that $Cf = \alpha Af + \beta Bf$ for $f \in \mathcal{D}(A) \cap \mathcal{D}(B)$. Then, $\{X_C(t), t \geq 0\}$ may be described as follows: conditional on being at $p \in \mathcal{S}$, $X_C(t)$ behaves like $X_A(t)$ or $X_B(t)$ with probabilities $\alpha(p)$ and $\beta(p)$, respectively.

In [10] we support the intuition presented above by showing that $\{T_C(t), t \geq 0\}$ may be approximated by the semigroups $\{\mathcal{T}_n(t), t \geq 0\}$ in $C_0(\mathcal{S}) \times C_0(\mathcal{S})$ generated by operators $\mathcal{A}_n$ with common domain $\mathcal{D}(\mathcal{A}_n) = \mathcal{D}(A) \times \mathcal{D}(B)$, given by $\mathcal{A}_n\binom{f}{g} = \binom{Af - n\beta f + n\beta g}{Bg + n\alpha f - n\alpha yg}$, for $n \geq 0$. More specifically, and the reader will not be surprised at the remark that this is yet another example of degenerate convergence of semigroups,

$$\lim_{n\to\infty} \mathcal{T}_n(t)\begin{pmatrix} f \\ g \end{pmatrix} = \begin{pmatrix} T_C(t)(\alpha f + \beta g) \\ T_C(t)(\alpha f + \beta g) \end{pmatrix}, \qquad f, g \in C_0(\mathcal{S}), t > 0.$$

Here $C_0(\mathcal{S}) \times C_0(\mathcal{S})$ (with the norm $\|\binom{f}{g}\| = \max\{\|f\|_{C_0(\mathcal{S})}, \|g\|_{C_0(\mathcal{S})}\}$) is identified via

isometric isomorphism with $C_0(\mathcal{S}_A \cup \mathcal{S}_B)$ where $\mathcal{S}_A = \mathcal{S} \times \{A\}$ and $\mathcal{S}_B = \mathcal{S} \times \{B\}$ are two copies of $\mathcal{S}$, and $\{\mathcal{T}_n(t), t \geq 0\}$ are viewed as Feller semigroups related to the following random processes $\{X_n(t), t \geq 0\}$ with values in $\mathcal{S}_A \cup \mathcal{S}_B$ (see Figure 1). While at $\mathcal{S}_A$, X_n behaves like X_A, in particular, it may jump with X_A, but stays in $\mathcal{S}_A$ until the random time τ when it leaves $\mathcal{S}_A$ to go to $\mathcal{S}_B$ and behave like X_B. At time τ, conditional on being at a point $(p, A) \in \mathcal{S}_A$, the process jumps to the copy (p, B) of this point in $\mathcal{S}_B$ and the intensity of jump varies with $p \in \mathcal{S}_A$, being equal to $n\beta(p)$. While at $\mathcal{S}_B$, the process behaves like X_B and the intensity of the jump to $\mathcal{S}_A$ is $n\alpha$.

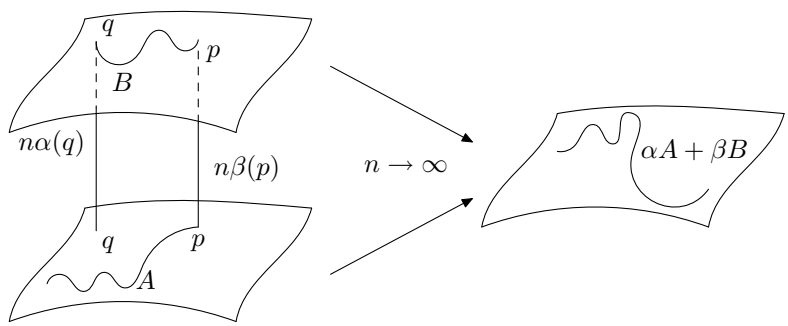


Fig. 1

It is helpful to think of X_n as a mixture of two processes. In the first of them, when starting at (p, A) we always stay at $\mathcal{S}_A$ and move according to the rules of process X_A; conditional on starting at (p, B) we stay at $\mathcal{S}_B$ all the time and move according to the rules of process X_B. In the second process the only possible transitions are jumps between $(p, A) \in \mathcal{S}_A$ and $(p, B) \in \mathcal{S}_B$. In other words, conditional on starting at (p, A) the second process is a Markov chain with two possible values (p, A) and (p, B), and with intensities $\begin{pmatrix} -n\beta(p) & n\beta(p) \\ n\alpha(p) & -n\alpha(p) \end{pmatrix}$ or, which is the same, with transition probabilities $\begin{pmatrix} \alpha(p) + \beta(p)\mathrm{e}^{-nt} & \beta(p) - \beta(p)\mathrm{e}^{-nt} \\ \alpha(p) - \alpha(p)\mathrm{e}^{-nt} & \alpha(p)\mathrm{e}^{-nt} + \beta(p) \end{pmatrix}, t \geq 0$. Hence, the second process, the process of jumps between $\mathcal{S}_A$ and $\mathcal{S}_B$, is composed of independent Markov chains indexed by p. As $n \to \infty$, for each $t > 0$, each of these chains reaches its equilibrium $(\alpha(p), \beta(p))$. Therefore, the limit process of X_n may indeed be thought of as the one behaving at p as X_A with probability $\alpha(p)$ or X_B with probability $\beta(p)$ (see Figure 1).

Apparently, there seems to be no other known approximation formula for the semigroup $\{T_C(t), t \geq 0\}$ that would support this intuition. In particular the Dyson-Phillips perturbation series and the Trotter product formula are of no help here (see [10] for more detailed discussion).

One particular example of a convex combination of two Feller generators is especially worth mentioning here. This is the case where $B = 0$ and α is separated from zero by a constant: $\alpha \geq \alpha_0 > 0$. By the main result of Dorroh [17] (which, by the way, is the precursor of the whole class of "multiplicative perturbation theorems" of semigroup theory – see [18]), αA generates a Feller semigroup. Moreover, B is the generator of the

semigroup composed of identities, which corresponds to no movement at all. Hence, in the light of our theorem, the process related to αA may be described as follows: conditional on being at $p \in \mathcal{S}$, with probability $\alpha(p)$ the process behaves like X_A and with probability $1 - \alpha(p)$ it stays at p. As a result, the process "slows down" as at each point it hesitates whether to move or to stay, and the smaller is α the slower the process moves. This agrees beautifully with the so-called Volkonskii formula [47] of probability theory, which says that multiplying the generator of a Feller process by a function is equivalent to a time-change in the process, a procedure of changing the speed at which the original proces runs through its path. In the case under consideration this function is $\alpha \leq 1$, and such a procedure results in "slowing down" of the process, in agreement with our interpretation.

Finally, let us mention that the resolvent of αA may be explicitly calculated in terms of the process X_A and this, via the use of the Feynman-Kac formula, leads to a straightforward proof of the (part of the) result of Dorroh and of the Volkonskii formula – see [10].

Having brushed against the Feynman-Kac formula, I am persuaded I came back home, to pure mathematics with well-grounded intuitions from applied sciences.

References

[1] G. K. Ackers, A. D. Johnson and M. A. Shea, *Quantitative model for gene regulation by λ phage repressor*, Proc. Natl. Acad. Sci. USA 79 (1982), 1129–1133.

[2] D. R. Akhmetov and R. Spigler, *Uniform and optimal estimates for solutions to singularly perturbed parabolic equations*, J. Evol. Equations 7 (2007), 347–372.

[3] W. Arendt, C. J. K. Batty, M. Hieber and F. Neubrander, *Vector-valued Laplace Transforms and Cauchy Problems*, Birkhäuser, 2001.

[4] O. Arino, E. Sánchez, R. Bravo De La Parra and P. Auger, *A singular perturbation in an age-structured population model*, SIAM J. Appl. Math. 50 (1999), 408–436.

[5] J. Banasiak, *Singularly perturbed linear and semilinear hyperbolic systems: kinetic theory approach to some folk's theorems*, Acta Applicandae Mathematicae 49 (1997), 199–228.

[6] A. Bobrowski, *Degenerate convergence of semigroups*, Semigroup Forum 49 (1994), 303–327.

[7] A. Bobrowski, *The Widder-Arendt theorem on inverting of the Laplace transform and its relationship with the theory of semigroups of operators*, Methods Funct. Anal. Topology 3 (1997), 1–39.

[8] A. Bobrowski, *Functional Analysis for Probability and Stochastic Processes*, Cambridge University Press, 2005.

[9] A. Bobrowski, *Degenerate convergence of semigroups related to a model of stochastic gene expression*, Semigroup Forum 73 (2006), 345–366.

[10] A. Bobrowski, *On a semigroup generated by a convex combination of two Feller generators*, J. Evol. Equ. 7 (2007), 555–565.

[11] A. Bobrowski, *On limitations and insufficiency of the Trotter–Kato theorem*, Semigroup Forum 75 (2007), 317–336.

[12] A. Bobrowski, T. Lipniacki, K. Pichór and R. Rudnicki, *Asymptotic behavior of distributions of mRNA and protein levels in a model of stochastic gene expression*, J. Math. Anal. Appl. 133 (2007), 753–769.

[13] A. Ciliberto, B. Novak and J. J. Tyson, *Steady states and oscilations in the p53/Mdm2 network,* Cell Cycle 4:3; March 2005, 488–493.
[14] M. H. A. Davis, *Piece-wise deterministic Markov processes*, J. Royal Statistical Soc. Ser. B 46 (1984) 353–388.
[15] M. H. A. Davis, *Lectures on Stochastic Control and Nonlinear Filtering*, Springer, 1984.
[16] M. H. A. Davis, *Markov Processes and Optimization*, Chapman and Hall, 1993.
[17] B. Dorroh, *Contraction semigroups in a function space*, Pacific J. Math. 19 (1966), 35–38.
[18] K.-J. Engel and R. Nagel, *One-parameter Semigroups for Linear Evolution Equations*, Springer, 2000.
[19] S. N. Ethier and T. G. Kurtz, *Markov Processes. Characterization and Convergence*, Wiley, 1986.
[20] W. Feller, *An Introduction to Probability Theory and Its Applications,* Vol. II, (second edition, 1971), Wiley, 1966.
[21] A. M. Femino, F. S. Fay, K. Fogarty and R. H. Singer, *Visualization of single RNA transcripts in situ*, Science 280 (1998), 585–590.
[22] J. A. Goldstein, *Semigroups of Linear Operators and Applications*, Oxford Mathematical Monographs, 1985.
[23] R. J. Griego and R. Hersh, *Random evolutions, Markov chains, and systems of partial differential equations*, Proc. Nat. Acad. Sci. USA 62 (1969), 305–308.
[24] R. J. Griego and R. Hersh, *Theory of random evolutions with applications to partial differential equations,* Trans. Amer. Math. Soc. 156 (1971), 405–418.
[25] B. Hat, P. Paszek, M. Kimmel, K. Piechór and T. Lipniacki, *How the number of alleles influences gene expression*, Journal of Statistical Physics 128 (2007), 511–533.
[26] E. Hille, *Functional Analysis and Semigroups*, Amer. Math. Soc., 1948.
[27] E. Hille and R. S. Phillips, *Functional Analysis and Semigroups*, Amer. Math. Soc., 1957.
[28] T. Hillen, *Hyperbolic models for chemosensitive movement*, Math. Models Methods of Appl. Sci. 12 (2002), no. 7 1–28.
[29] K. Ito and H. P. Mc Kean, Jr., *Diffusion Processes and Their Sample Paths*, Springer, 1965.
[30] M. Kac, *Some Stochastic Problems in Physics and Mechanics*, Magnolia Petroleum Co. Colloq. Lect. 2, 1956.
[31] T. B. Kepler and T. C. Elston, *Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations*, Biophysical J. 81 (2001), 3116–3136.
[32] J. F. C. Kingman, *The coalescent*, Stoch. Proc. Appl. 13 (1982), 235–248.
[33] J. Kisyński, *On second order Cauchy's problem in a Banach space*, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astr. et Phys. 18 (1970), 371–374.
[34] J. Kisyński, *On cosine operator functions and one-parameter groups of operators*, Studia Math. 44 (1972), 93–105.
[35] T. G. Kurtz, *Extensions of Trotter's operator semigroup approximation theorems,* J. Funct. Anal. 3 (1969), 354–375.
[36] A. Lasota and M. C. Mackey, *Chaos, Fractals and Noise*, Springer, 1994.
[37] T. Lipniacki, P. Paszek, A. R. Brasier, B. Luxon and M. Kimmel, *Mathematical model of NF-kB regulatory module*, J. Theor. Biol. 228 (2004): 195–215.
[38] T. Lipniacki, P. Paszek, A. Marciniak-Czochra, A. R. Brasier and M. Kimmel, *Transcriptional stochasticity in gene expression,* J. Theor. Biol. 238 (2006), 348–367.
[39] T. Lipniacki, P. Paszek, A. R. Brasier, B. Luxon and M. Kimmel, *Stochastic regulation in early immune response*, Biophys. J. 90 (2006), 725–742.

[40] M. Lisi and S. Totaro, *The Chapman–Enskog procedure for an age-structured population model: initial, boundary and corner layer corrections*, Mathematical Biosciences 196 (2005), 153–186.
[41] P. Mandl, *Analytical Treatment of One-Dimensional Markov Processes*, Springer, 1968.
[42] R. Nagel (ed.), *One-Parameter Semigroups of Positive Operators*, Springer, 1986.
[43] P. Paszek, *Modeling stochasticity in gene regulation: characterization in the terms of the underlying distribution function*, Bulletin of Mathematical Biology 69 (2007), 1567–1601.
[44] A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations*, Springer, 1983.
[45] A. Raj, C. S. Peskin, D. Tranchina, D.Y. Vargas and S. Tyagi, *Stochastic mRNA synthesis in mammalian cells,* PLoS Biol 4 (10) (2006): e309. DOI: 10.1371/journal.pbio.0040309.
[46] D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*, third edition, Springer, 1999.
[47] L. C. G. Rogers and D. Williams, *Diffusions, Markov Processes and Martingales, Vol. 1, Foundations*, Cambridge University Press, 2000.
[48] M. A. Shea and G. K. Ackers, *The Or control system of bacteriophage lambda: a physical-chemical model for gene regulation*, J. Mol. Biol. 181 (1985), 211–230.
[49] M. Sova, *Convergence d'opérations linéaires non bornées*, Rev. Roum. Math. Pures et App. 12 (1967), 373–389.
[50] M. Thattai and A. van Oudenaarden, *Intrinsic noise in gene regulatory networks*, Proc. Natl. Acad. Sci. USA 98 (2001), 8614–8619.
[51] J. R. R. Tolkien, *The Hobbit, or There and Back Again*, Bellantine Books, 1965.
[52] K. Yosida, *Functional Analysis*, Springer, 1965.

STOCHASTIC MODELS IN BIOLOGICAL SCIENCES
BANACH CENTER PUBLICATIONS, VOLUME 80
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WARSZAWA 2008

GENERIC PRINCIPLES OF ACTIVE TRANSPORT

MAURO MOBILIA, TOBIAS REICHENBACH, HAUKE HINSCH,
THOMAS FRANOSCH and ERWIN FREY

Arnold Sommerfeld Center for Theoretical Physics (ASC) and Center for NanoScience (CeNS)
Department of Physics, Ludwig-Maximilians-Universität München
Theresienstrasse 37, D-80333 München, Germany
E-mail: mauro.mobilia@physik.uni-muenchen.de, tobias.reichenbach@physik.lmu.de,
hinsch@physik.uni-muenchen.de, franosch@physik.uni-muenchen.de, frey@lmu.de

Abstract. Nonequilibrium collective motion is ubiquitous in nature and often results in a rich collection of intriguing phenomena, such as the formation of shocks or patterns, subdiffusive kinetics, traffic jams, and nonequilibrium phase transitions. These stochastic many-body features characterize transport processes in biology, soft condensed matter and, possibly, also in nanoscience. Inspired by these applications, a wide class of lattice-gas models has recently been considered. Building on the celebrated *totally asymmetric simple exclusion process* (TASEP) and a generalization accounting for the exchanges with a reservoir, we discuss the qualitative and quantitative nonequilibrium properties of these model systems. We specifically analyze the case of a dimeric lattice gas, the transport in the presence of pointwise disorder and along coupled tracks.

1. Introduction. The modeling of collective transport phenomena is an important challenge in theoretical physics, with possible technological and interdisciplinary implications (e.g. in biology, soft matter and nanotechnology). In this context, motion of molecular motors [How], single-file diffusion in colloid supsensions in confined geometry [Lin], spintronic devices [Zut], have motivated the study of a class of simple one-dimensional stochastic models. Remarkably, and despite their apparent simplicity, these models have been shown to be characterized by rich physical properties like nonequilibrium phase transitions, mixed phases, traffic jams, which are of interest on their own right. Hence, the goal of this article is to convey a pedagogical review of these lattice-gas models and of the techniques which allow to unravel their intriguing features.

2000 *Mathematics Subject Classification*: Primary 82C70; Secondary 82C26, 92C45.

Key words and phrases: lattice-gas model, totally asymmetric simple exclusion process.

M.M. gratefully acknowledges the support of the German Alexander von Humboldt Foundation through the Fellowship No. IV-SCZ/1119205 STP. The authors are grateful to R. Kouyos, A. Parmeggiani and P. Pierobon for pleasant collaboration on some topics reported in this work.

The paper is in final form and no version of it will be published elsewhere.

 © Instytut Matematyczny PAN, 2008

For the case of biological engines, it is only recently that modern experimental techniques [Meh] have revealed the detailed causes of sub-cellular motion and transport. Nowadays, we know that every use of our muscles is the collective effort of a class of proteins called myosin that "walk" on actin filaments. All proteins that convert the chemical energy of ATP (adenosine-triphospate) in a hydrolysis reaction into mechanical work are referred to as molecular motors. These biological "engines" are highly specialized in their tasks: ribosomes move along mRNA strands while translating the codons into proteins, dynein is responsible for cilia motion and axonal transport, and kinesins play a key role in cytoskeletal traffic and spindle formation [How]. Rather than on the exact details of the molecular structure and function of motor proteins (see e.g. [Schl]), we are interested in phenomena arising out of the collective interaction of many motors. Early research along this line was motivated by mRNA translation that is managed by ribosomes, which bind to the mRNA strand with one subunit and step forward codon by codon. To increase the protein synthesis many ribosomes can be bound to the same mRNA strand simultaneously. This fact might induce collective properties, the slow down of an incoming ribosome (and of the protein synthesis) due to steric hindrance caused by another ribosome in front, as was first realized by MacDonald [Mac], who set up a theoretical model for the translation of highly expressed mRNA.

Much in the spirit of [Mac] our theoretical analysis focuses on simple, yet essential, collective processes underlying the motion and neglects the chemical or mechanical details on the molecular level of motor steps. Within this approach, even simple stochastic models are found to exhibit rich collective behavior.

We will start this article by revisiting (Section 2) the main properties of the celebrated totally asymmetric simple exclusion process (TASEP). In Section 3, we review the main properties of a recently introduced model where the TASEP is coupled to a process accounting for the exchanges with a reservoir. In the following sections (Sec. 4-7), we outline recent results on variants of such a model and therefore discuss its robustness. In particular, the case of a lattice gas of extended particles is considered in Sec. 4, while the effects of pointwise disorder are addressed in Sec. 5. Sections 6 and 7 are devoted to the transport properties along two types of coupled tracks. In Section 8, we conclude this work by providing an outlook and our conclusions.

2. The TASEP. Biological engines, like ribosomes and kinesins, often move along one-dimensional tracks (e.g. mRNA strands or microtubules). Each individual "step" is the result of a fascinating sequence of molecular events and is an interesting research topic on its own right [How], but is not our scope here. In fact, our approach is to devise simple and analytically amenable stochastic models. By identifying and mimicking the main ingredients underlying the collective transport, our goal is to get collective phenomena which we expect to be (qualitatively) comparable to those observed.

A first step along this line of research was done in a pioneering work by MacDonald et al. [Mac], where the the total asymmetric exclusion process (TASEP) was introduced. This has since been widely recognized as a paradigmatic model in nonequilibrium statistical mechanics. It consists of a one-dimensional lattice (Fig. 1, left) with N sites labeled

by $i = 1, \cdots, N$ and with a spacing of $a = L/N$. Here, L is the total length of the lattice and, for technical convenience, is often set to 1 with the lattice spacing then referred to as $\varepsilon = 1/N$.

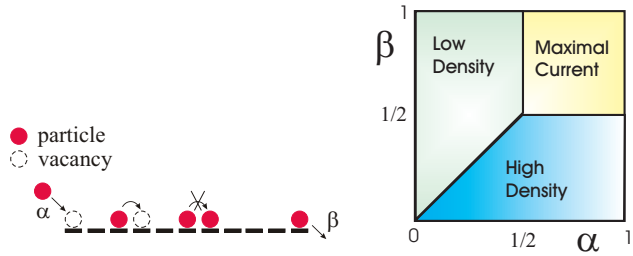


Fig. 1 (Color online). (Left) Schematic model of TASEP (particles are injected with rate α, move exclusively to the right being subject to hard-core exclusion, and are removed with rate β). (Right) Phase diagram of TASEP in α, β phase space. One distinguishes the low density (LD, green), the high density (HD, blue) and the maximal current (MC, yellow) phases, as well as the first ($\alpha = \beta < 1/2$) and second order transition lines (in bold).

Particles (mimicking the motors) have an extension of the size of the lattice spacing and are subjected to hard core exclusion due to steric hindrance. Along the track, particles can jump to their right neighboring site with rate r (set to 1 in the following) provided the latter is empty, whence the occupation number n_i of site i takes either the value 0 or 1. In addition, one has to specify the dynamic rules at the boundaries of the lattice. Contrary to equilibrium systems, the latter have a crucial influence and are responsible for *boundary-induced phase transitions* [Kru]. While one can consider periodic or reflective boundaries, the richest behavior for the TASEP is obtained in the presence of open boundaries: at the left boundary ($i = 1$) particles attempt to enter the system with a rate α, while they exit the lattice at the right boundary ($i = N$) with rate β. This is equivalent to two additional sites $i = 0$ and $i = N+1$ at the boundaries, which are connected to the system by the bulk dynamics described above, and are constantly set to the density α and $1 - \beta$ respectively. Because it does not fulfill the detailed balance, the TASEP evolves into a steady state where a non-vanishing current is maintained between the boundaries. Upon varying these boundary conditions, the TASEP was found to exhibit phase transitions which – following general theorems [Mer] – are not even allowed for one-dimensional equilibrium systems in the absence of long-range interactions.

2.1. *Analysis of the TASEP.* While a severe difficulty for the analysis of non-equilibrium systems stems from the lack of universal concepts like the Boltzmann-Gibbs ensemble, models like the TASEP and its generalization can still be studied in many details by means of methods which are discussed in the remainder of this review article. Here, for the sake of clarity, we outline the analysis of the TASEP and some of its main statistical properties. The latter are related to the density and current distributions, which are quantities of direct biological relevance.

At any given moment, the system can be found in a certain configuration $\mathcal{C}$ made up of the occupation numbers at each lattice site. After a stochastic event occurred, i.e. the jump of one particle to a neighboring site, the new configuration is $\mathcal{C}'$. By its very

definition, the TASEP is a Markov process. Hence, the probability of finding the system in the configuration $\mathcal{C}$ at time t obeys the the master equation:

$$\frac{\mathrm{d}}{\mathrm{d}t}P(\mathcal{C};t) = \sum_{\mathcal{C}'\neq\mathcal{C}} \left[\omega_{\mathcal{C}'\to\mathcal{C}}P(\mathcal{C}';t) - \omega_{\mathcal{C}\to\mathcal{C}'}P(\mathcal{C};t)\right] , \tag{1}$$

where the $\omega_{\mathcal{C}\to\mathcal{C}'}$ is the transition rate from the configuration $\mathcal{C}$ to $\mathcal{C}'$. Using, for instance, the so-called "quantum Hamiltonian formalism" [Sch], one can derive from (1) the equations of motion for the density and current of particles. In fact the density of particle at site i, $\langle n_i\rangle$, is found to obey

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\langle n_i\rangle = \langle n_{i-1}(1-n_i)\rangle - \langle n_i(1-n_{i+1})\rangle, \quad (i=2,\ldots,N-1),\\
&\frac{\mathrm{d}}{\mathrm{d}t}\langle n_1\rangle = \alpha\left[1-\langle n_1\rangle\right] - \langle n_1(1-n_2)\rangle,\\
&\frac{\mathrm{d}}{\mathrm{d}t}\langle n_N\rangle = \langle n_{N-1}(1-n_N)\rangle - \beta\langle n_N\rangle,
\end{aligned} \tag{2}$$

where the brackets stand for the average over the histories and, for brevity, we have omitted the time-dependence. From the discrete continuity equation (with $\nabla j_i \equiv j_i - j_{i-1}$), $\frac{\mathrm{d}}{\mathrm{d}t}\langle n_i\rangle + \nabla j_i = 0$, one also infers the expression for the local current of particles:

$$j_i = \langle n_i(1-n_{i+1})\rangle. \tag{3}$$

The interpretation of (2, 3) is straightforward, but the analysis is difficult because these equations underlie an infinite hierarchy of equations: the density is connected to the two-point correlations, which depend on the three-point and so on. Despite these difficulties, an exact solution has been achieved by using the so-called matrix and Bethe Ansatz methods [Der1,Der2,Sch]. As these exact methods are mostly tailored for the TASEP, it is insightful to outline an alternative approach. The latter relies on a mean-field assumption and, while it is approximate in its essence, it has a broad range of application and allows to derive the main properties of the models discussed in this review.

We now focus on the stationary properties of the TASEP and show how the phase diagram can be obtained within a simple *mean-field theory*. The latter relies on neglecting all the spatial correlations: $\langle \hat{n}_i\hat{n}_j\rangle = \varrho_i\varrho_j$, where we have introduced $\varrho_i \equiv \langle n_i\rangle$. In the steady state ($d\varrho_i(t)/dt = 0$), this approximation leads to the following equations of motion for the density in the bulk:

$$0 = \varrho_{i-1}(1-\varrho_i) - \varrho_i(1-\varrho_{i+1}), \tag{4}$$

while the expression for the current reads

$$j_i = \varrho_i(1-\varrho_{i+1}). \tag{5}$$

Clearly, the hierarchy of equations is now closed and, together with boundary conditions, forms a nonlinear set of equation giving rise to nontrivial solutions. A crucial observation is the local conservation of the stationary current through the bulk. In fact, from Eq. (4) it follows that $j_i = j_{i-1}$. Actually, one remarks that such a relation is also embodied in Eq. (2) and holds true beyond the mean-field approximation. At this point, it is useful to adopt a continuum limit which turns the spatial lattice variable continuous. We thus consider a large number N of lattice sites and rescale the length of the system to $L = 1$,

with infinitesimal lattice space $\varepsilon \equiv L/N = 1/N \ll 1$. Introducing the spatial variable $x = i/N$, with $0 \leq x \leq 1$, one can proceed with a Taylor expansion of the local density: $\varrho(x \pm \varepsilon) = \varrho(x) \pm \varepsilon \partial_x \varrho(x) + \frac{1}{2}\varepsilon^2 \partial_x^2 \varrho(x) + O(\varepsilon^3)$. In the continuum limit and by neglecting the higher order terms in ε, Eq. (4) is recast as a simple first-order differential equation:

$$(2\varrho - 1)\partial_x \varrho = 0. \tag{6}$$

The latter has to be supplemented with the boundary conditions $\varrho(0) = \alpha$ and $\varrho(1) = 1-\beta$. Because we have a first order differential equation that needs to satisfy two boundary conditions, we are evidently concerned with an over-determined boundary value problem. There are three solutions to (6): $\varrho_{\rm bulk}(x) = 1/2$ does not satisfy either boundary condition (except for the special case $\alpha = \beta = 1/2$), while $\varrho(x) = C$ can satisfy either the left or the right boundary condition, resulting in $\varrho_\alpha(x) = \alpha$ and $\varrho_\beta(x) = 1 - \beta$, respectively. To obtain a solution satisfying the boundary conditions, both solutions need to be matched. To resolve this issue, both solutions have to be considered valid in non-overlapping areas of the system. At the point $x_{\rm w}$ where the current imposed by the left and right boundaries coincides (conservation of the current), those areas border and are connected by a sharp domain wall, such that $\varrho = \varrho_\alpha$ for $0 \leq x \leq x_{\rm w}$ and $\varrho = \varrho_\beta$ for $0 \leq x_{\rm w} \leq 1$. The domain wall is found to move with a velocity $V = \beta - \alpha$ [Kol1]. These simple considerations, together with the conservation of the current, have important implications and result in three distinct phases:

- For $\beta < 1/2$ and $\alpha > \beta$, the current $j_\alpha = \varrho_\alpha(1 - \varrho_\alpha)$ of the left density solution ϱ_α is higher than the current in the right part of the system and the particles are transported faster to the DW from the left end, then they head on the right. Hence, the domain wall is shifted to the left and the whole bulk density takes the value of ϱ_β (except for a small boundary layer, which vanishes in the limit $N \to \infty$). This situation therefore corresponds to a *high density* phase (HD), where both the density and the current are determined by the exit rate β: the mean-field theory gives a spatially constant density $\rho_i = 1 - \beta$ larger than $1/2$ and a constant current $j_i = \beta(1 - \beta)$. In this case, the current is dominated by the low exit rate which acts as a bottleneck for the transport.
- Along the same lines, for $\alpha < 1/2$ and $\alpha < \beta$ the low entrance rate is the limiting factor for the particle current which is now given by $j_i = \alpha(1 - \alpha)$. Since $\rho_i = \alpha$ is always smaller than $1/2$ in this parameter range, the phase is also termed the *low density* (LD) phase.
- If both α and β become larger than the critical value $1/2$, the current saturates and the density becomes constant $\rho^* = 1/2$ independently of the parameters at the boundaries. The current is limited by the particle exclusion in the bulk and its maximal value is $j^* = 1/4$; therefore, this phase was termed *maximal current* (MC) phase.

The (first-order) transition line $\alpha = \beta < 1/2$ separating the LD and HD phases requires special treatment. There, the velocity of the domain wall is zero, but Monte Carlo simulations and rigorous results [And] show that the DW actually performs a random walk in a domain with reflecting boundaries and an average over a sufficiently

long time will therefore result in a homogeneous probability density over space (instead of a localized shock) $\varrho(x) = \alpha + (1 - \beta - \alpha)x$.

The above discussion is summarized in the phase diagram of the TASEP (Fig. 1), where the LD and HD phases are separated by a first-order transition line, while the HD/MC and LD/MC phases are separated by a second-order transition line. It has to be noticed that the phase diagram predicted by the mean-field theory coincides with that obtained by exact methods. Actually, the only features not accurately accounted for by the mean-field theory are the boundary layers and the density profile along the line $\alpha = \beta < 1/2$ [Der1,Der2,Sch]. The somewhat suprising validity of the mean-field approach can be traced back to the density relation, $j = \varrho(1 - \varrho)$, which has been shown to be an exact result.

3. The TASEP with on-off kinetics. With the aim to cover more realistic situations, there has recently been an upsurge of lattice gas models in the spirit of the TASEP. In fact, to devise more realistic transport models, some of the simplistic assumptions underlying the TASEP had to be questioned. As an example, experimental observations have shown that ribosomes typically cover an area on the mRNA that exceeds the unit lattice space. To account for this situation the particles in TASEP have to extend over several lattice sites. This feature was accounted for by considering the ℓ-TASEP, where extended particles (of length ℓ) move unidirectionally being subject of excluded volume, and it was found that the TASEP phase diagram was not affected qualitatively [Sha]. It was also shown that the presence of clusters of bottlenecks, mimicking slow codons along messenger RNA, may either give rise to local density perturbations or affect macroscopically the density profile [Chou,Tri].

Intracellular transport along cytoskeletal filaments has also served as a source of inspiration for driven lattice gas models. While in the TASEP model motors can only bind and unbind on the left and the right boundary respectively, cytoskeletal motors are known to detach from the track to the cytoplasm [How], where they perform Brownian motion, and subsequently reattach to the track. The interplay between diffusion in the cytoplasm and directed motion along the filament was studied [Lip] both in open and closed compartments, focusing on anomalous drift and diffusion behavior, as well as on maximal current and traffic jams as function of the motor density. In Ref. [Par1] it has been realized that, for an appropriate scaling of the on-off rates that ensures that particles always travel a finite fraction on the lattice, the on-off kinetics may not only give rise to quantitative changes in the transport efficiency but also to a novel class of driven lattice gas models. Actually, it was shown that the interplay between bulk on-off kinetics and driven transport results in a steady state exhibiting phase separation. In the following subsection, we review the main results of the studies [Par1,Par2].

3.1. *The TASEP/LK.* The crucial feature of cytoskeletal transport is the possibility of bulk attachment and detachment with a finite residence time on the lattice. The latter can be understood as an effect of thermal fluctuations that may overcome the binding energy of the motors that is only of the order of several $k_{\mathrm{B}}T$. In this respect, attachment and detachment of motors is a stochastic process whose dynamic rules have to

be defined. Parmeggiani *et al.* [Par1] chose to use Langmuir kinetics (LK), known as adsorption-desorption kinetics of particles on one- or two-dimensional lattices coupled to a bulk reservoir [Vil1]. Particles can adsorb at empty sites and desorb from occupied sites, with microscopic reversibility demanding that the kinetic rates obey detailed balance leading to an evolution towards an equilibrium steady state (in the statistical mechanics sense). In this framework, the choice of LK is especially appealing as one is then faced with the competition of a prototypic equilibrium (Langmuir kinetics) and a paradigmatic non-equilibrium (TASEP) system. The resulting system – in the following referred to as TASEP/LK – is hence defined as follows: the well-known TASEP is extended with the possibility of particles to attach to the filament with rate ω_{A} and to detach from an occupied lattice site to the reservoir with rate ω_{D}. According to the type of ensemble (canonical or grand canonical) the reservoir is either finite or infinite. Here, the reservoir is assumed to be infinite and homogeneous throughout space and time. For a lattice experiencing only LK, the equilibrium coverage depends solely on the ratio $K = \omega_{\mathrm{D}}/\omega_{\mathrm{A}}$, the resulting steady state, given by Langmuir isotherm $\varrho_{\mathrm{L}} = K/(K+1)$, is completely uncorrelated for neglecting any particle interaction (except excluded volume). This appears sensible as the diffusion in the cytoplasm is fast enough to flatten any deviations from the homogeneous reservoir density. If we now consider the combination

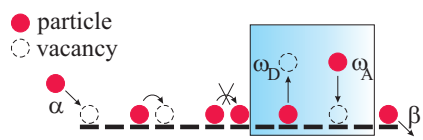


Fig. 2. Schematic model of TASEP/LK: the TASEP is extended by possible particle attachment and detachment in the bulk with rate $\omega_{\mathrm{A}}, \omega_{\mathrm{D}}$

of TASEP and LK into the model displayed in Fig. 2, due attention has to be paid to the different statistical nature of the processes. On the one hand, TASEP evolves into a non-equilibrium state carrying a finite current and, the number of particles being conserved through the bulk, the system is very sensitive to the boundary conditions. On the other hand, LK as an equilibrium process is expected to be robust against any boundary effects (especially for large systems). Combining both processes on the same time scale would unavoidably lead to a trivial domination of LK, as the bulk rates ω_{A} and ω_{D} apply to a large number of bulk sites and would become predominant over the local entrance and exit rates α and β. To observe any interesting behavior (i.e. real interplay) between the two dynamics, one needs to ensure competition in the system. A prerequisite for the two processes to compete is to deal with comparable reaction rates, which requires an appropriate scaling. To this end N-independent global detachment and attachment rates, Ω_{D} and Ω_{A}, are introduced, while the local on-off rates per site scale as

$$\omega_{\mathrm{D}} = \frac{\Omega_{\mathrm{D}}}{N};\ \omega_{\mathrm{A}} = \frac{\Omega_{\mathrm{A}}}{N}. \tag{7}$$

The physical meaning of the above scaling is revealed by focusing on the involved time scales: a particle on the lattice performs a move on a time scale which is the inverse of the move's rate. Therefore, a particle spends an average time $\approx 1/\omega_{\mathrm{D}}$ on the lattice before

detaching. Bearing in mind that the TASEP jump rate is set to unity, a particle jumps to its right neighboring site typically after one unit time step. Therefore, the particle will travel a number $N_{\rm T} = 1/\omega_{\rm D}$ of sites before leaving the lattice. Compared to the lattice length this corresponds to a fraction of $n_{\rm T} = N_{\rm T}/N = 1/(N\omega_{\rm D})$. To keep this fraction finite, and ensure effective competition between transport and on-off dynamics, $\omega_{\rm D}$ needs to scale as defined in (7). Concretely, for the kinesin the ratio between (un)binding and diffusion rates is typically $10^{-1} - 50^{-1}$ [How].

3.2. *Mean field solution of TASEP/LK.* To obtain density and current distributions of the TASEP/LK, we use again a mean-field approach. Hence, by neglecting any spatial correlations, along the same lines of Sec. 2.1, the stationary density is found to obey

$$0 = \varrho_{i-1}(1-\varrho_i) - \varrho_i(1-\varrho_{i+1}) - \omega_{\rm D}\varrho_i + \omega_{\rm A}(1-\varrho_i), \tag{8}$$

which has to be supplemented with the same boundary conditions as above. It is again useful to consider the continuum limit, where (at first order in ϵ) the above equation reduces to

$$(2\varrho - 1)\partial_x \varrho - \omega_{\rm D}\varrho + \omega_{\rm A}(1-\varrho) = 0, \tag{9}$$

with the boundary conditions $\varrho(0) = \alpha$ and $\varrho(1) = 1 - \beta$. To proceed with the analysis of Eq. (9), it is useful to introduce the binding constant $K \equiv \omega_{\rm A}/\omega_{\rm D}$ between the attachment and detachment rates. However, for the sake of clarity, here we restrict ourselves to the simple situation with the same attachment and detachment rates, i.e. $K = 1$, and refer the readers to Ref. [Par2] for a comprehensive discussion of the mathematically more involved case where $K \neq 1$. For equal on and off rates, $\Omega_A = \Omega_D = \Omega$, it is enlightening to rewrite Eq. (9) as $(\partial_x \varrho - \Omega)(2\varrho - 1) = 0$. Obviously, there are two solutions to this nonlinear differential equation: the homogeneous density $\varrho_{\rm L} = 1/2$, given by the Langmuir isotherm, and the linear slope $\varrho(x) = \Omega x + C$, the constant C being determined by the boundary conditions, which leads to one solution imposed by the entrance boundary, $\varrho_\alpha(x) = \alpha + \Omega x$, and another one, namely $\varrho_\beta(x) = 1 - \beta - \Omega + \Omega x$, determined by the exit end. The complete density profile $\varrho(x)$ is the combination of one or several pieces of these three functions. Depending on how these are matched, we are led to distinguish several phases as explained below.

3.3. *Phase diagram and density distributions.* The only region in the parameter space of the TASEP/LK which is the same as for the simple TASEP is associated with fast entrance and exit rates. In this case, the system is again in a maximal current phase, i.e. in a bulk-controlled regime, and the upper right quadrant of the phase diagram (Fig. 3, right) do not change compared to the phase diagram of the TASEP (Fig. 1). In this regime, the additional bulk on-off dynamics at Langmuir isotherm $\varrho_{\rm L} = 1/2$ do not result in any changes in the density distribution. In this case, non-equilibrium and equilibrium dynamics do not compete but cooperate.

As in the TASEP, different solutions can be matched in various ways, the simplest being the connection by a domain wall between the left solution ϱ_α and the right solution ϱ_β. Depending on the current distribution, two possibilities have to be distinguished. As both solutions are spatially inhomogeneous, the corresponding currents j_α and j_β are strictly monotonic (Fig. 3, left). If the currents equal each other inside the system at a position

x_w, the domain wall is localized at this position, as a displacement to either side would result in a current inequality driving the DW back to x_w (see Fig. 3, left). As a consequence, the TASEP/LK displays multi-phase coexistence of low and high density regions (LD-HD phase) in the stationary state on any time scales. This has to be contrasted with the TASEP, where this behavior is only predicted for short observation times. Recently, this domain wall localization has been claimed to be observed experimentally [Nish].

If the matching of left and right currents is not possible inside the system, the known LD and HD phases are found. This is the case for one boundary condition being considerably larger than the other and depends on the slope of the density profile. This slope is determined by the ratio of the TASEP step rate and the bulk interchange rate Ω: For large Ω any density imposed by the boundary relaxes fast against the Langmuir isotherm of $\varrho_L = 1/2$, resulting in a steep slope of the density profile.

This fact allows for the existence of two other phases with multi-regime coexistence. We could imagine a scenario in which the boundary-imposed density solutions decay fast enough towards the isotherm to enable a three-regime coexistence of low density, maximal current and high density (LD-MC-HD phase). Furthermore, a combination of a MC phase with a boundary layer on one side and a LD or HD region of finite extend at the other boundary can be imagined. Not all these phases will be realized for every value of Ω. Instead, the phase topology of two-dimensional cuts through the α, β, Ω-phase space changes. An example is shown in Fig. 3 (right).

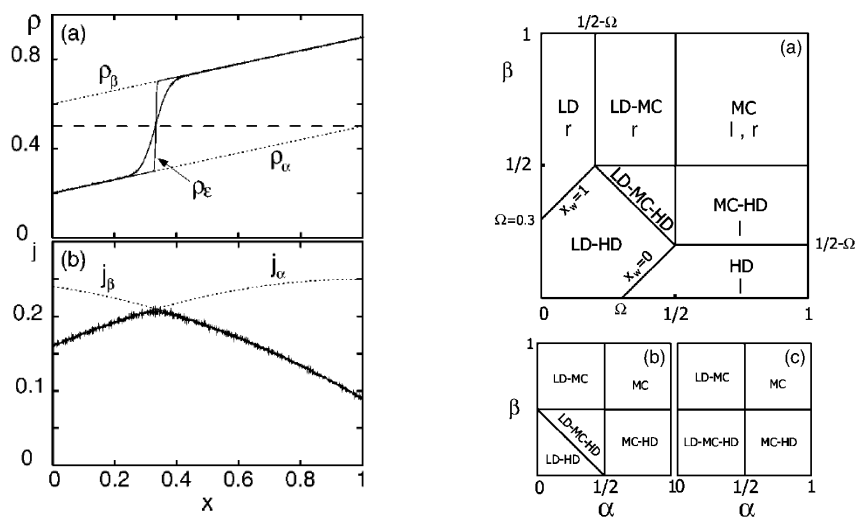


Fig. 3 (Reproduced from [Par2]). (Left) The DW connects the two densities ρ_α and ρ_β (*both dashed*) and is localized at the point where the correspondent currents j_α and j_β match. Here low and high densities are connected together through a domain wall, and the resulting phase is termed LD-HD. Note the finite extend of the domain wall (localization length) that is only produced for Monte-Carlo simulations (*solid wiggly line*) and is not captured by mean-field results (*solid line*), reproduced from [Hin1]. (Right) Topological changes in the phase diagram of TASEP/LK for (a) $\Omega = 0.3$, (b) $\Omega = 0.5$, (c) $\Omega = 1$.

4. Dimers and robustness. The results discussed in the previous sections have been obtained within the framework of generic models. The key notion is that a detailed description of a physical system is nonessential for the occurrence of the phenomenon, but a small number of ingredients encodes a mechanism that leads to the phenomenon. In order to corroborate that the models studied so far are representative for a wide class of models it is instructive to make a step towards reality and add new features to the model.

In the case of molecular motors, their molecular structure already encodes how the motion of a single motors is performed. The most studied examples are kinesins, dyneins and myosin V which are known to consist of two sub-components referred to as *lead* and *trail head*. These heads bind specifically to a particular subunit of a long polar molecule acting as a molecular track. As the dynamics of ribosomes along the mRNA motivated the study of $\ell-$TASEP for the transport of molecular motors, it was also suggested that a realistic minimal extension of the TASEP/LK should comprise *dimers*, i.e. objects that occupy two lattice sites at any instant of time. These extended objects move according to the following elementary processes, depicted in Fig. 4, which underlie a generalization of the above TASEP/LK dynamics:

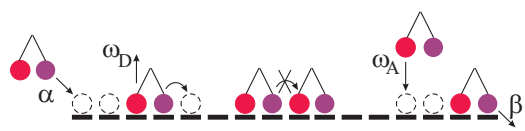


Fig. 4 (Color online). Illustration of the model of driven dimer transport coupled to a reservoir

Unidirectional hopping: If the lead head of a dimer occupies a site and the following is empty, the dimer advances one step to the right with unit rate.

Entrance: At rate α, a dimer enters the lattice with its trail head provided that the first two sites are empty.

Exit: A dimer with its lead head on the last site leaves the lattice with rate β emptying the last two sites.

Detachment: Everywhere in the bulk (i.e. trail head on a site $i = 2, \ldots, N-2$) a dimer leaves the lattice emptying two sites with a site independent detachment rate ω_D.

Attachment: Everywhere in the bulk a dimer enters the lattice with its trail head, provided that the considered site as well as its right neighbor is empty, with a site-independent attachment rate ω_A.

Although at first glance it appears to be a straightforward extension of the monomer model, a theoretical description encounters a number of pitfalls. First, the mean-field approach based on ignoring any correlations between neighboring sites is too simplistic even on the level of pure on-off kinetics. There, the fact that two neighboring sites have to be empty to allow for an attachment process modifies the relation of the equilibrium density $\rho_{\rm L}$, i.e. the Langmuir isotherm, and the binding constant $K = \omega_A/\omega_D$. Due to the hard-core exclusion, the corresponding statistical mechanics calculation reduces to a combinatorial problem [Tonk,Thom] yielding the thermodynamic equation of state $\rho_I = \rho_{\rm L}(K)$ [McG].

The consequences of extended objects on the stationary density profiles for a *pure* TASEP have been studied in [Lak, Shaw]. In particular, the authors of those works found that the phase diagram is in essence the one of monomeric TASEP, the differences being of quantitative nature only. A theoretical description has been achieved within a 'refined mean-field theory' that accounts approximately for the correlation induced by simultaneous hops of two monomers rigidly glued together. Special care had to be taken to link the density at the boundary to the entrance/exit rates.

Recently, it has been recognized that the same refined mean-field ansatz is the proper starting point also to deal with the combined problem of driven transport (TASEP) and on-off kinetics [Pier1]. The Langmuir isotherm ρ_{L} as well as an appropriate current-density relation is encoded in this approach. The density profile in the stationary state again results from a first order differential equation subject to two (non-trivial) boundary conditions. In particular, the scenario of phase separation and phase coexistence has been established for that model, and it has been rationalized that branch points and domain walls are generic features of such one-dimensional driven transport models. An important conclusion of these studies is that the phenomena that emerge upon weakly coupling a bulk process to a boundary controlled system are robust.

Stochastic simulations have quantitatively corroborated the phase diagram obtained analytically. In particular, as a sensitive test of the theory the position of the domain wall has been measured and perfect agreement with the theoretical prediction has been found. Again it appears that a mean-field type approach is able to capture the phase diagram exactly, as well as the shape of the stationary density profiles. The weak coupling to the reservoir is sufficient to wash out the dynamic correlations that build up as dimers walk through the system. Similarly, the slow time scale for the on-off kinetics that intrinsicly would lead to the random sequential adsorption is sped up by the continuous mass transport through the system.

5. Pointwise disorder. The presence of some form of randomness on the track is another realistic ingredient which is highly desirable to incorporate in a model for intracellular transport. In fact, *disorder* may biologically be mediated by structural imperfections of the microtubular structure or proteins associated with the microtubules that change the affinity of the motors with the track. In the TASEP/LK language, imperfections on the microtubule can be mimicked by the presence of *pointwise disorder* [Chou]. In addition, studying the effect of disorder on dynamic and steady state properties of nonequilibrium systems is a major challenge in statistical physics [Stin].

In this framework, we have studied the influence of a bottleneck on the stationary transport properties of the TASEP/LK which is arguably its simplest, yet nontrivial, disordered version. Namely, as in the previous sections, we have considered the competition between the totally asymmetric exclusion process and Langmuir kinetics in the mesoscopic limit [see Sec. 3.1, Eq. (7)] and in the presence of open boundaries, see Fig. 2. However, the system is now supplemented with the presence of a bottleneck of strength $q < 1$, which, at a given site k, slows down any incoming particles (Fig. 5, top). Depending on the current of particles through the system and the strength of the defect, it is

clear that the latter may have either a local (spike) or macroscopic effect (jump in the density profile) [Pier2].

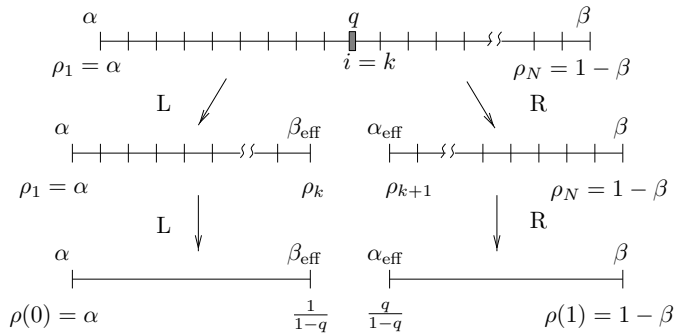


Fig. 5 (Reproduced from [Pier2]). Schematic illustration of the effective mean-field theory: At the defect's location, $i = k$, the lattice is split into two subsystems (L and R), which are connected through the effective rates α_{eff} and β_{eff}. The last step represents the continuum limit. The density $\rho(x_d)$ at the bottleneck's location is $1/(1-q)$ on L and $q/(1-q)$ on R.

In the region of the parameter space where the bottleneck is macroscopically relevant, the properties of the system are investigated through an *effective mean-field theory* in the continuum limit ($k/N \to x_d$). As illustrated in Fig. 5, this approach is built on splitting the lattice into two subsystems on the left and right of the defect (respectively termed L and R). To ensure the local conservation of the current (which holds also in the presence of a bottleneck), one has to couple the L and R subsystems through the effective entrance and exit rates $\alpha_{\text{eff}} = q/(1+q)$ and $\beta_{\text{eff}} = 1/(1+q)$ on the right and left ends of L and R (Fig. 5). Therefore, one is left with two coupled TASEP/LK processes whose current profiles have still to be matched. As the current has a spatial dependence (due to the Langmuir kinetics), the defect depletes its profile within a distance ξ, that we call *screening length*. As the latter quantity increases with the strength of the defect and decreases with the attachment-detachment rates, it is an important concept for our analysis. In fact, depending on the screening length and the position of the defect, one has to distinguish between various scenarios, which are efficiently characterized by the corresponding *carrying capacity*. This notion stands for the maximal local current that can flow through the bulk of the entire system (including the defect).

For the sake of simplicity, here we focus on the situation with equal attachment and detachment rates (i.e. $\Omega_D = \Omega_A = \Omega$). As the screening length either covers entirely the system or part of it, we identify four types of carrying capacities $\mathcal{C}_i(x), i = 1, \ldots, 4$. Using the concepts of screening length and carrying capacity, one readily obtains the phase diagram of the TASEP/LK system perturbed by a localized bottleneck. In fact, when the boundary currents dominate, the phase behavior of the defect-free system is recovered. Also, above some critical entrance and exit rates, the system transports the maximal current (carrying-capacity of the system), which is independent of the boundary rates. Between these two extreme situations, we find several coexistence phases: above some specific parameter values the system is dominated by the defect and the phase diagram

is characterized by new *bottleneck phases* (BP). When the screening length spans the whole system, the associated carrying capacity, say $\mathcal{C}_1$, never reaches the maximal value $j^* = 1/4$ and one finds *four new mixed phases*. Among the latter, three phases (termed LD-BP, BP-HD and LD-BP-HD, see [Pier2]) are characterized by localized shocks in the density profile. When the screening length is short and the related carrying capacity, say $\mathcal{C}_2$, saturates at value j^* within both subsystems L and R, one identifies *nine novel bottleneck-induced phases*. When the screening length is asymmetric (carrying capacities $\mathcal{C}_3$ and $\mathcal{C}_4$), i.e. covers entirely the subsystem R (L) and partially L (R), there are *six new mixed phases* [Pier2].

Each of these new mixed phases has been quantitatively studied within our mean-field theory and the analytical results have been checked against numerical simulations. As an illustration, let us consider the phase corresponding to the density profile of Fig. 6. In this case, the screening length is short and the carrying capacity is of type $\mathcal{C}_2$. As shown in Fig. 6 (right), near the extremities of the track, the current is imposed by the boundaries ($j = j_\alpha$ and $j = j_\beta$) and is below the maximal value j^*, which is reached within the bulk. This gives rise to a *low* and *high density* (sub-)phases. At sufficient distance from the ends and from the defect, the current saturates at its maximal value j^*, which corresponds to a *maximal current* (sub-)phase. Within a distance ξ from the defect (screening length; Fig. 6, right), the latter dictates the current profile and the system is in a *bottleneck phase*. As a consequence, the density profile displayed in Fig. 6 (left) corresponds to a phase termed LD-MC-BP-MC-HD.

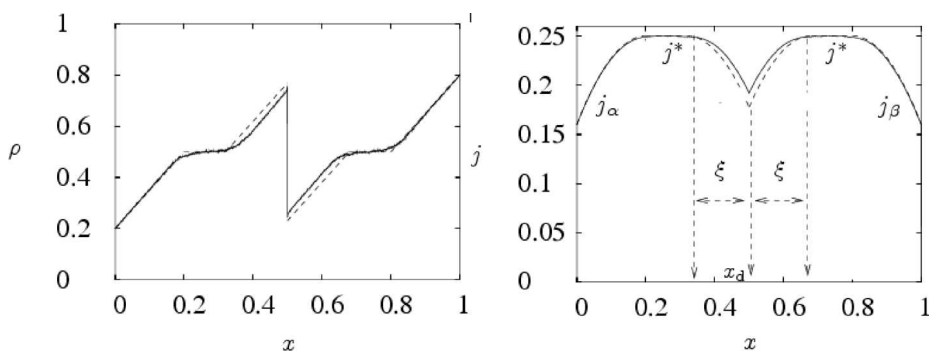


Fig. 6. Example of the TASEP/LK density (left) and current (right) profiles in the presence of a bottleneck located at $x_d = 1/2$. The system (with parameters $q = 0.3$, $\alpha = \beta = 0.2$, $\Omega_D = \Omega_A = 1.5$ and $N = 4096$) is in a mixed phase (LD-MC-BP-MC-HD), where (from the left to the right end) a low density (LD) and maximal current (MC) subphases are followed by a bottleneck-induced (sub)phase (BP). The latter is also concatenated to a MC and a high density (HD) subphases (see text). Results of simulations (continuous line) are compared to analytic predictions (dashed line).

While the simple TASEP is known to be not affected qualitatively (only some transitions lines are shifted in the phase diagram) by the presence of a localized defect [Jan,Kol2], the effective competition between TASEP and LK turns out to be very sensitive to the presence of pointwise disorder. We expect the same holds for other disordered

versions of the TASEP/LK, e.g. in the presence of clusters of bottlenecks [Chou], to which the above effective theory can be adapted.

6. Multiple parallel lanes. Microtubules, the intracellular tracks for molecular motors like dynein or kinesin [How], are built of 12-14 parallel lanes. Although it has been revealed that motor proteins typically remain on one track while proceeding on the microtubule, the statistics of deviations (random lane changes) is so far unknown. In the following, we aim to investigate effects of a small, but finite number of lane changes on the resulting density profiles along the lanes by studying a transport model consisting of two parallel lanes I and II, see Fig. 7 and [Rei1, Rei2].

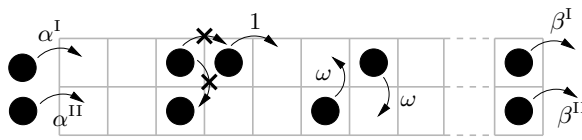


Fig. 7 (Reproduced from [Rei1]). Illustration of a two-lane exclusion process. Lanes I and II possess individual entering rates, α^{I} resp. α^{II} as well as exiting rates, β^{I} resp. β^{II}.

Models with indirect coupling of lanes or strong direct coupling through frequent lane changes have been investigated in [Pop1, Pop2, Pro, Mit], while the focus of the work presented in the following is on weak coupling emerging from only a few lane changes. Particles are injected at the left boundary at rates α^{I} (α^{II}) on lane I (II), provided the respective lattice site is empty. In the bulk, particles hop forward at a constant rate which we set to unity. In addition, they may change from lane I to II and back at a small rate ω (weak lane coupling), always under the constraint of simple site exclusion. Having reached the right boundary, particle extraction occurs at rates β^{I} (β^{II}).

Interestingly, the above introduced model allows for an interpretation in terms of spin transport. Indeed, on a given lattice site, a particle may be on lane I or II, which may be viewed as an internal state, e.g. the spin state of an electron. Simple site exclusion translates into Pauli's exclusion principle: spin-up and spin-down may simultaneously occupy the same site, though not two particles of the same spin state. The resulting dynamics, depicted in Fig. 8, may have possible applications within the field of spintronics [Zut].

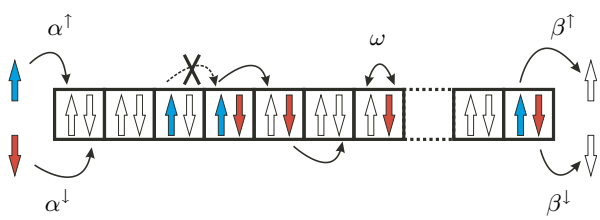


Fig. 8 (Color online; reproduced from [Rei1]). The same model as above, in the spin-transport interpretation. Particles respect Pauli's exclusion principle; spin flips occur at a (small) rate ω.

As for the problem of transport with attachment and detachment of hard-core particles in the bulk [Sec. 3.1, Eq. (7)], we employ a mesoscopic scaling for the exchange rate ω

by keeping the gross lane change rate, defined as $\Omega = \omega L$, as constant for large system size L. The latter yields a measure for the total number of lane changes that a particle undergoes while traversing the system, which is thus kept fixed, in this way ensuring a competition of lane change events with the boundary processes (injection and extraction).

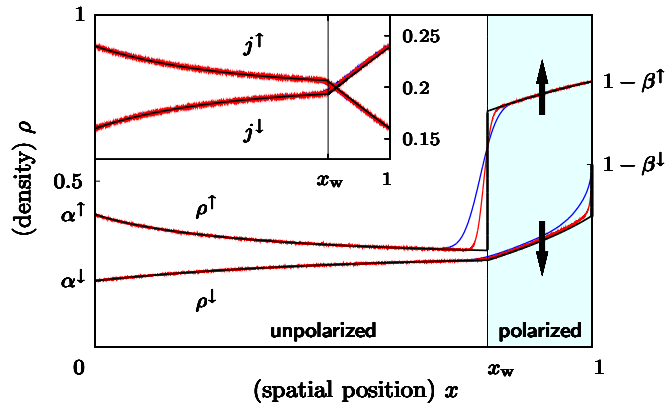


Fig. 9 (Reproduced from [Rei1]). Typical scenario for the non-equilibrium stationary state: a shock forms in the spin-up density profile, inducing a spontaneous polarization. The spin currents, shown in the inset, are both continuous. Solid lines correspond to the analytical solution, while dashed lines indicate results from stochastic simulations on lattices with $L = 2000$ (blue or dark grey) and $L = 10000$ (red or light gray).

This model exhibits a variety of different phases for the emerging non-equilibrium steady state, controllable by the injection and extraction rates as well as the lane change rate. A typical scenario is presented in Fig. 9, where the language of spin transport has been adopted: $\rho^{\uparrow}$ denotes the average density of spin-up (i.e. the concentration of particles on lane I) and $\rho^{\downarrow}$ stands for the density profile of spin-down (lane II) particles. While the density of spin-down remains at low values, the density of spin-up undergoes a transition at x_w from a region of low to a region of high value with the emergence of a localized shock. In the spin transport interpretation, we encounter a spontaneous polarization: the densities of both spin states rather equal each other to the left of the shock, but largely differ to the right, yielding a polarization there. A key point for the analytical description is the continuity (though not smoothness) at x_w of the individual spin currents, shown in the inset.

In the same spirit of the previous sections, we have taken advantage of a mean-field approach and a continuum limit to analytically describe the stationary state density profiles. Surprisingly, comparison with stochastic simulations (finite size scaling) uncovers the apparent exactness of the analytical results in the limit of large system sizes. Furthermore, our analytical approach allows for the derivation of phase diagrams, leading to the full knowledge of the system's phase behavior.

7. SEP and TASEP coupled on a ring. While not only in the TASEP but also in all extended models presented above, the strategy of a mean-field approach and the

continuum limit has proven succesful to obtain correct phase diagrams and density profiles in the limit of large system sizes, this cannot a priori be taken for granted. This was recently demonstrated in a study on the competitive effects of driven versus diffusive motion in lattice gases [Hin2]. As prominent examples we consider the two paradigms of driven and diffusive transport in lattice gases: the purely driven TASEP, as introduced above, and the purely diffusive symmetric exclusion process (SEP). The analyzed model systems is a periodic setup (Fig. 10, left) consisting of a SEP and a TASEP part of equal size N with N_p particles inside the system. The hopping rate of unity on the lower lane signifies a choice of time units and the hopping rate D on the upper lane fixes the ratio between driven and diffusive motion.

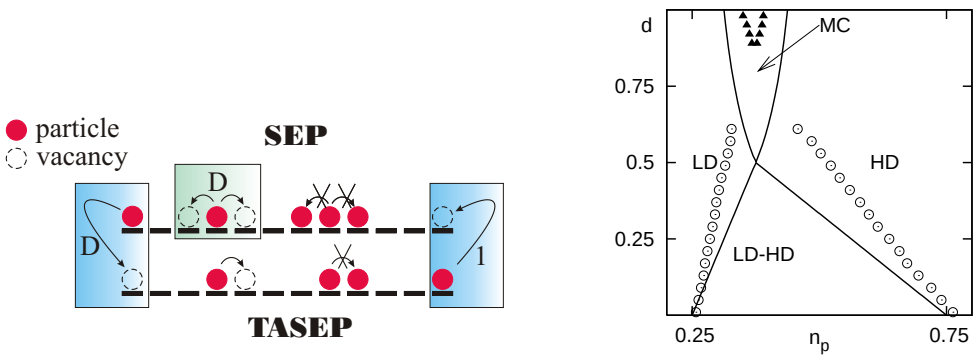


Fig. 10 (Color online). (Left) Schematic illustration of the ring system. The dynamics of the upper lane is governed by a SEP with rate D, and the dynamics of the lower lane by a TASEP with unity jump rate. The boundary densities are indicated at the junctions. (Right) Reproduced from [Hin2]. Phase diagram obtained by MF (solid lines) exhibits four phases. Simulations [MC (triangles) and LD-HD (circles) phase boundaries] reveal a failure of the MF analysis.

In a first step, we proceed along the lines of analysis used above to derive a mean-field phase diagram. Here, the central idea is to decouple the two parts of the ring and consider them as separate lanes with effective entrance and exit rates. Known results for both processes can then be applied. While TASEP's non-equilibrium phase transitions have been introduced in Sec. 2, SEP is known to exhibit a linear density profile connecting the boundary values [Sch]. The periodicity of the system has two important consequences: current conservation and conservation of particle density $n_p = N_p/2N$. The current conservation allows one to establish the relations $\alpha = D\delta$ and $\gamma = 1 - \beta$ between the boundary densities. The system's state can then be described by the two remaining control parameters D and n_p. The non-equilibrium steady states have been classified and a phase diagram is derived as a function of those two parameters.

In the case of TASEP/LK a mesoscopic scaling was used to ensure competition for arbitrary large systems. To enlight the behavior of the ring system, the current is of crucial importance. While the lower lane carries a current $J_L = \rho(1-\rho)$ independent of system size, the upper lane's current is determined by the boundary condition's difference as $J_U = (\gamma - \delta)D/N$ reminding of "Fick's law". Obviously, this current decreases with the

number of sites, leading to a "frozen" system due to an unphysical vanishing of current. However, a mesoscopic scaling $d = D/N$ ensures a competition between both processes and a finite current as expected in the case of continuous diffusion in the limit $N \to \infty$.

Using a mean-field analysis and the conservation laws, phase transitions can be identified at characteristic changes in the TASEP density profile. As for the simple TASEP a LD and a HD phase are observed. However, a crucial difference is the existence of a localized domain wall that is observed in simulations and whose position x_w can be derived as a function of the new control parameters as $x_w = (-3 + 3d + 4n_p)/(4d - 2)$. There exists thus a LD-HD phase of finite extend in the ring system. This can be explained by particle conservation and is a consequence of the periodic system being canonical instead of the grand canonical TASEP. Additionally, the variance of the domain wall position is found to decay much faster with N^{-1} compared with a $N^{-1/2}$ decay in TASEP. The extend of the LD-HD phase is computed to (see Fig. 10, right):

$$\begin{aligned} d &= \frac{3 - 4n_\mathrm{p}}{3} \quad \text{for} \quad x_\mathrm{w} = 0, \\ d &= 4n_\mathrm{p} - 1 \quad \text{for} \quad x_\mathrm{w} = 1. \end{aligned} \tag{10}$$

By the introduction of a finite LD-HD phase, the first-order phase transition between LD and HD phase is substituted by two second order transitions. Similar to TASEP a MC phase can be identified with phase boundaries:

$$d = \frac{1}{16n_\mathrm{p} - 4} \qquad \text{and} \qquad d = \frac{1}{8 - 16n_\mathrm{p}}\,. \tag{11}$$

In a second step, we used Monte Carlo simulations to validate the derived phase diagram. It is found that strong deviations (symbols in Fig. 10 right) occur. Thinking of the success of the mean-field method in similar systems, this comes as a surprise at first sight. Investigating its origin, a closer look at the density profile reveals that the upper part of the density profile is actually strictly convex instead of linear. This is a result of a strong time scale separation of the two subprocesses introduced by the mesoscopic scaling. While diffusive hopping events happen on a very short time scale, the entrance of a particle at the r.h.s. junction is a rather rare event. Between these entrance events the particle surplus at the r.h.s. relaxes diffusively to the left. As a result the time averaged density profile appears convex and "Fick's law" for the current is no longer valid. The current is effectively diminished since it is now bounded to its lowest value at the point of smallest local gradient. This explains the qualitative shift of the critical point in the phase diagram to higher d.

The observed behavior bears analogies to forest fire models [Dro], where out of two processes of different time scales the faster one strives towards equilibrium and the slower one occasionally forces the system to a non-equilibrium state. It furthermore reminds of the coexistence line in simple TASEP with open boundaries where a linear profile is obtained as the time average over a randomly walking domain wall. Hence, the study of the ring system highlights the importance of time scale separation and points out the limits of mean-field theory in lattice gas models.

8. Outlook and conclusion. In this work, we have described how complex nonequilibrium transport phenomena, e.g. the motion of molecular motors along microtubuli or transport of spins, have inspired stochastic models displaying rich nonequilibrium collective properties. Namely, in the same spirit as mRNA translation was modeled by means of the *totally asymmetric simple exclusion process* (TASEP) [Mac], intracellular transport processes have motivated the study of a broad class of lattice gas models. To account for exchanges with a reservoir (e.g. mimicking the binding/unbinding of motors from/to the cytosol) and the track, TASEP was supplemented with an on-off (Langmuir) kinetics. An effective competition between driven motion and Langmuir kinetics arises provided the related attachment and detachment rates scale with the inverse of the system size. In this case, a finite residence time on the lattice is ensured and results in the emergence of a new class of non-equilibrium features, like multi-phase coexistence, localized shocks and enriched phase diagrams. Along these lines, we have reported recent results for physically motivated variants of the above model.

Specifically, we have discussed transport properties of a one-dimensional gas of dimers, which was motivated by the fact that molecular motors like kinesin have two heads. The finite extension of the particles then has led to a breaking of particle/hole symmetry accompanied by a failure of a naïve mean-field approach. A more sophisticated approach had to be developed that allows to calculate the exact phase diagram (essentially) analytically. It has been argued that the scenario of phase separation is generic for this kind of models. We have also considered the effects of pointwise disorder in its simplest form, i.e. in the presence of a localized *bottleneck* slowing down any incoming particle. By means of an effective mean-field theory and numerical simulations, we have shown that such a single defect is responsible for the emergence of numerous new mixed phases and can result in the simultaneous formation of two traffic jams. The presence of multiple parallel lanes on which motors proceed can lead to further interesting phenomena. We have studied a two-lane model with rare lane switch events of the particles and discussed the mapping to a possible spintronic device. Depending on boundary rates, the stationary density profiles can exhibit jammed states appearing on one or both lanes. Finally, the study of competition between driven and diffusive motion on two lanes in a periodic setup has shown the limitations of mean-field theory in the presence of strong time-scale separation.

References

[And] E. D. Andjel, M. D. Bramson, and T. M. Liggett, *Shocks in the asymmetric exclusion process*, Probab. Th. Rel. Fields 78 (1988), 231.

[Chou] T. Chou and G. Lakatos, *Clustered bottlenecks in mRNA translation and protein synthesis*, Phys. Rev. Lett. 93 (2004), 198101.

[Der1] B. Derrida, E. Domany and D. Mukamel, *An exact solution of a one-dimensional asymmetric exclusion model with open boundaries*, J. Stat. Phys. 69 (1992), 667.

[Der2] B. Derrida and M. R. Evans, *Exact correlation-functions in an asymmetric exclusion model with open boundaries*, J. Physique I 3 (1993), 311.

[Dro] B. Drossel and F. Schwabl, *Self-organized critical forest-fire model*, Phys. Rev. Lett. 69 (1992), 1629.

[Gar] C. W. Gardiner, *Handbook of Stochastic Methods*, Springer, Berlin, 1983.

[Hin1] H. Hinsch, R. Kouyos and E. Frey, *From intracellular traffic to a novel class of driven lattice gas models*, in: Traffic and Granular Flow '05, A. Schadschneider et al. (eds.), Springer, Berlin, 2006.

[Hin2] H. Hinsch and E. Frey, *Bulk-driven nonequilibrium phase transitions in a mesoscopic ring*, Phys. Rev. Lett. 97 (2006), 095701.

[How] J. Howard, *Mechanics of Motor Proteins and the Cytoskeleton*, Sinauer Press, Sunderland, 2001.

[Jan] S. A. Janowsky and J. L. Lebowitz, Phys. Rev. A 45 (1992), 618.

[Kol1] A. B. Kolomeisky, G. M. Schütz, E. B. Kolomeisky and J. P. Straley, *Phase diagram of one-dimensional driven lattice gases with open boundaries*, J. Phys. A 31 (1998), 6911.

[Kol2] A. B. Kolomeisky, *Asymmetric simple exclusion model with local inhomogeneity*, J. Phys. A 31 (1998), 1152.

[Kru] J. Krug, *Boundary-induced phase transitions in driven diffusive systems*, Phys. Rev. Lett. 67 (1991), 1882.

[Lak] G. Lakatos and T. Chou, *Totally asymmetric exclusion processes with particles of arbitrary size*, J. Phys. A: Math. Gen. 36 (2003), 2027.

[Lin] B. Lin, M. Meron, B. Cui, and S. A. Rice, *From random walk to single-file diffusion*, Phys. Rev. Lett. 94 (2005), 216001.

[Lip] R. Lipowsky, S. Klumpp and T. M. Nieuwenhuizen, *Random walks of cytoskeletal motors in open and closed compartments*, Phys. Rev. Lett. 87 (2001), 108101.

[Mac] C. T. MacDonald, J. H. Gibbs and A. C. Pipkin, *Kinetics of biopolymerization on nucleic acid templates*, Biopolymers 6 (1968), 1.

[McG] J. McGhee and P. von Hippel, *Theoretical aspects of DNA-protein interactions: Co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice*, J. Mol. Biol. 86 (1974), 469.

[Meh] A. D. Mehta, M. Rief, J. A. Spudich, D. A. Smith and R. M. Simmons, *Single-molecule biomechanics with optical methods*, Science 283 (1999), 1689.

[Mer] N. D. Mermin and H. Wagner, *Absence of ferromagnetism or antiferromagnetism in one- or two-dimensional isotropic Heisenberg models*, Phys. Rev. Lett. 17 (1966), 1133.

[Mit] T. Mitsudo and H. Hayakawa, *Synchronisation of kinks in the two-lane totally asymmetric simple exclusion process with open boundary conditions*, J. Phys. A: Math. Gen. 38 (2005), 3087.

[Nish] K. Nishinari, Y. Okada, A. Schadschneider and D. Chowdhury, *Intracellular transport of single-headed molecular motors KIF1A*, Phys. Rev. Lett 95 (2005), 118101.

[Par1] A. Parmeggiani, T. Franosch and E. Frey, *Phase coexistence in driven one dimensional transport*, Phys. Rev. Lett. 90 (2003), 086601.

[Par2] A. Parmeggiani, T. Franosch and E. Frey, *Totally asymmetric simple exclusion process with Langmuir kinetics*, Phys. Rev. E 70 (2003), 046101.

[Pier1] P. Pierobon, T. Franosch and E. Frey, *Driven lattice gas of dimers coupled to a bulk reservoir*, Phys. Rev. E 74 (2006), 031920.

[Pier2] P. Pierobon, M. Mobilia, R. Kouyos and E. Frey, *Bottleneck-induced transitions in a minimal model for intracellular transport*, Phys. Rev. E 74 (2006), 031906.

[Pop1] V. Popkov and I. Peschel, *Symmetry breaking and phase coexistence in a driven diffusive two-channel system*, Phys. Rev. E 64 (2001), 026126.

[Pop2] V. Popkov and G. M. Schütz, *Shocks and excitation dynamics in a driven diffusive two-channel system*, J. Stat. Phys. 112 (2003), 523.

[Pro] E. Pronina and A. B. Kolomeisky, *Asymmetric coupling in two-channel simple exclusion processes*, Physica A 372 (2006), 12.

[Rei1] T. Reichenbach, T. Franosch and E. Frey, *Exclusion processes with internal states*, Phys. Rev. Lett. 97 (2006), 050603.

[Rei2] T. Reichenbach, E. Frey and T. Franosch, *Traffic jams induced by rare switching events in two-lane transport*, New J. Phys. 9 (2007), 159.

[Sch] G. Schütz, in: Phase Transitions and Critical Phenomena, C. Domb and J. Lebowitz (eds.), Academic Press, London, 2000.

[Schl] M. Schliwa and G. Woehlke, *Molecular motors*, Nature 422 (2003), 759.

[Sha] L. B. Shaw, P. K. P. Zia and K. H. Lee, *Totally asymmetric exclusion process with extended objects: A model for protein synthesis*, Phys. Rev. E 68 (2003), 021910.

[Stin] R. B. Stinchcombe, *Disorder in non-equilibrium models*, J. Phys.: Condens. Matter 14 (2002), 1473.

[Thom] C. Thompson, *Classical Equilibrium Statistical Mechanics*, Oxford University Press, New York, 1988.

[Tonk] L. Tonks, *The complete equation of state of one, two and three-dimensional gases of hard elastic spheres*, Phys. Rev. 50 (1936), 955.

[Tri] G. Tripathy and M. Barma, *Driven lattice gases with quenched disorder: Exact results and different macroscopic regimes*, Phys. Rev. E 58 (1998), 1911.

[Vil1] E. Frey and A. Vilfan, *Anomalous relaxation kinetics of biological lattice-ligand binding models*, Chem. Phys. 284 (2002), 287.

[Van] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, North-Holland, Amsterdam, 1981.

[Zut] I. Žutić, J. Fabian, and S. Das Sarma, *Spintronics: Fundamentals and applications*, Rev. Mod. Phys. 76 (2004), 323-410.

STOCHASTIC MODELS IN BIOLOGICAL SCIENCES
BANACH CENTER PUBLICATIONS, VOLUME 80
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WARSZAWA 2008

DRIFT, DRAFT AND STRUCTURE: SOME MATHEMATICAL MODELS OF EVOLUTION

ALISON M. ETHERIDGE

Department of Statistics, University of Oxford, UK
E-mail: etheridg@stats.ox.ac.uk

Abstract. Understanding the evolution of individuals which live in a structured and fluctuating environment is of central importance in mathematical population genetics. Here we outline some of the mathematical challenges arising from modelling structured populations, primarily focussing on the interplay between forwards in time models for the evolution of the population and backwards in time models for the genealogical trees relating individuals in a sample from that population. In addition to classical models we describe a special case of a new model introduced in very recent work with Nick Barton. A number of directions for future research are suggested.

1. Introduction. If we are to make inferences from genetic data then we must understand how genes evolve in a structured and fluctuating population. Two types of structure are important. First, individuals live in a particular spatial location and their rate of reproduction depends on the local environment and competition from other individuals living nearby. Second, genes are embedded in different *genetic* backgrounds—because genes are organised on chromosomes and chromosomes in turn are grouped into individuals, different genes do not evolve independently of one another. In particular, the evolution of a gene which is itself selectively neutral in that it does not confer any particular advantage or disadvantage to the organism that carries it can nonetheless be influenced by natural selection acting on other genes in that same organism.

Modelling structured populations presents challenging mathematical problems and our aim here is to outline some of the existing approaches and their drawbacks. Of particular importance is the interplay between forwards in time models for the evolution of a population and backwards in time models for the *ancestry* of a population. We shall mostly be concerned with spatial structure rather than genetic structure, but we shall

2000 *Mathematics Subject Classification*: Primary 60K35, 92D10, 92D15, 92D40; Secondary 60G52, 60G55, 60G57.

Key words and phrases: spatial population models, extinction-recolonisation, coalescents.

Research supported in part by EPSRC grants GR/T19537/01, GR/S41203/01

The paper is in final form and no version of it will be published elsewhere.

 © Instytut Matematyczny PAN, 2008

argue that some of the mathematics that arises in the study of populations undergoing selection could also throw new light on populations evolving in a spatial continuum and in this context we will describe a new model which overcomes some of the shortcomings of existing models. However, there remain many more questions than answers.

In §2 we describe Kingman's coalescent model for the ancestry of an idealised population and introduce the notion of genetic drift. In a *neutral* theory of evolution in which all individuals in the population are equally fit, it is this genetic drift, which corresponds to random reproduction in the population, that drives the changes in frequency of different genetic types. But in such a theory this basic model predicts that genetic variability in a population is determined by the size of the population and this is simply not consistent with data. What we observe is low variation in polymorphism levels between species in contrast to the high variation in census population sizes. Moreover, a theory of evolution based on genetic drift alone predicts dramatically more genetic variability than observed in nature. Other evolutionary forces must be incorporated into our model. In §3 we consider the effect of natural selection on genealogies. In particular we discuss *genetic hitchhiking* and present a model due to Gillespie in which the effect of genetic drift is supplemented by what he dubs genetic *draft*. Under Gillespie's model the ancestry of the population is described by a coalescent with multiple mergers. We recall some basic results about such coalescents in §4. In §5 we turn to *spatial* structure and outline the stepping stone model and some of its shortcomings. The most obvious difficulty is that it supposes the population to be *subdivided* and does not easily extend to the case where it is not. In §6 we present a new model, very recently introduced in joint work with Nick Barton, which not only applies to populations evolving in a spatial continuum but also allows for the inclusion of certain large-scale extinction-recolonisation events which have dominated the demographic history of many species. Under this model, the ancestry of the population is described by what one can call a *spatial* Λ-coalescent or more generally a spatial Ξ-coalescent. It is more general than other processes of that name and we believe biologically more relevant. Our new model incorporates demography through extinction-recolonisation events, but it treats them as instantaneous and it does not, at this stage, allow for fluctuations in population size or interactions between competing species. We discuss some demographic models in §7 before, finally, in §8 suggesting some directions for future research.

REMARK 1.1 (Some remarks on terminology). We shall use interchangeably the terms *gene* and *genetic locus*. For our purposes it is convenient to think of chromosomes as linear and genes or genetic loci being points on that linear set. The different forms that a gene can take, which we will use to label individuals in the population, are called *alleles*.

2. Genetic drift. Typically the patterns in genetic data are used to make inferences about the genealogical relationships between individuals in a sample from a population and so it is these genealogical relationships, in other words the family trees that relate individuals in the sample, that we try to model.

The most basic model is the classical Kingman coalescent. This process was introduced by Kingman (1982) and it provides a simple and elegant description of the genealogical re-

lationships between individuals in a *large*, randomly mating population of constant size in which all individuals are equally fit. We're going to suppose for simplicity that we are modelling a *haploid* population which just means that each individual has exactly one parent.

REMARK 2.1 (Diploid populations). For diploid populations such as our own, in which chromosomes are carried in pairs, provided one is only interested in one gene one typically treats the chromosomes as if they formed a haploid population of size $2N$, ignoring the pairing into individuals. Because, as we explain in §3, in diploid populations chromosomes are not typically passed down as indivisible blocks from parent to offspring, this device does not work if one is following more than one genetic locus. In that setting the genetic types at different loci on a chromosome may be derived from two different chromosomes in the previous generation and to trace the full ancestry we must follow both parental lineages.

Before we can describe the genealogical trees, we need a model for the way in which the population evolves *forwards* in time. The workhorse of mathematical population genetics is the Wright-Fisher model. In this model the population evolves in discrete generations. The family sizes of individuals in the parental population are determined by multinomial sampling with equal weights on all individuals. An equivalent (and convenient) way to say that is the following:

DEFINITION 2.2 (The neutral Wright-Fisher model). A population of N genes evolves in discrete generations. Generation $(n+1)$ is formed from generation n by choosing N genes at random with replacement. i.e. Each gene in generation $(n+1)$ chooses its parent uniformly at random from those present in generation t.

Suppose now that we take a sample of size k from such a population. It is an easy matter to determine the family trees which relate individuals in the sample. For example, suppose that $k = 2$. The probability that the two individuals share a common parent is $1/N$. Conditional on having distinct parents, the probability that they share a common grandparent is again $1/N$ and so on. Thus the time to the most recent common ancestor of the two individuals has a geometric distribution with success parameter $1/N$. This has mean N and so for large populations it is convenient to measure time in units of N generations. The time to the most recent common ancestor of our sample of size two is then approximately exponentially distributed with parameter 1.

More generally, for a sample of size k, the probability that three or more of them share a common parent is $\mathcal{O}(1/N^2)$ and the chance that two *different* pairs of individuals share common parents is also $\mathcal{O}(1/N^2)$. For large N then the time until we see *any* event as we trace back the ancestral lineages is approximately the minimum of $\binom{k}{2}$ exponentially distributed random variables each with parameter one, in other words an exponentially distributed time with parameter $\binom{k}{2}$. At that time it is equally likely to be any of the pairs of ancestral lineages which *coalesce* into a single lineage. We then wait an exponentially distributed time with parameter $\binom{k-1}{2}$ until the next event—which is a coalescence of two of the surviving lineages, again chosen at random—and so on. For large populations this then provides an approximate description of the genealogy of the sample if we measure time in units of size N.

Notice that because we are modelling a haploid population (so individuals have only one parent) the family tree gets *smaller* as we go backwards in time, unlike the growing trees that describe your family history in which you track two parents of every ancestor (cf. Remark 2.1).

If we label individuals in the population $\{1, \ldots, k\}$ then the process we have just described induces a continuous time Markov chain on the space of equivalence relations on $[k] = \{1, 2, \ldots, k\}$. Individuals with common ancestor at time t before the present form a single equivalence class.

DEFINITION 2.3 (Kingman's coalescent). A $[k]$-*coalescent* is a continuous time Markov chain on $\mathcal{E}_k$, the space of equivalence relations on $[k]$, with transition rates $q_{\xi,\eta}$ $(\xi, \eta \in \mathcal{E}_k)$ given by

$$q_{\xi,\eta} = \begin{cases} 1 & \text{if } \eta \text{ is obtained by coalescing two of the equivalence classes of } \xi, \\ 0 & \text{otherwise.} \end{cases}$$

The *Kingman coalescent* on $\mathbb{N}$ is a process of equivalence relations on $\mathbb{N}$ with the property that, for each k, its restriction to $[k]$ is a k-coalescent.

We shall use the term coalescent for both the process of coalescing ancestral lineages and the induced partition-valued process.

In passing to the large population limit we have taken a *diffusion approximation*. The rôle of the Wright-Fisher model is to identify the *correct* diffusion approximation. We would have obtained the same approximate genealogical trees for a wide variety of finite population models. In particular, provided that the variance of the number of offspring of each individual remains bounded as we pass to the large population limit, (up to a constant time change reflecting this variance) Kingman's coalescent provides an approximation to the genealogical trees for any exchangeable population model. It is this robustness of Kingman's coalescent which makes it such a powerful tool.

To see why one might call this a diffusion approximation, consider the effect of letting population size tend to infinity on the forwards in time population model. For simplicity suppose that the population is subdivided into two genetic types which we label a, A. We continue to suppose that these types are selectively neutral so that all individuals in our population are equally fit. When we pass to the infinite population limit we will not be able to talk about the *numbers* of individuals of each type, but we will be able to talk about the *proportion* of each type at time t (measured in units of size N). Let us write p_t for the proportion of the population at time t which is of type a. Then in a single generation, corresponding to a time interval of length $\delta t = 1/N$, since the number of a individuals in the new generation has the binomial distribution with N trials and success probability p, the change Δp in that proportion satisfies

$$\mathbb{E}[\Delta p] = 0, \quad \mathbb{E}[(\Delta p)^2] = \frac{1}{N^2} Np(1-p) = \delta t p(1-p), \tag{1}$$

and

$$\mathbb{E}[(\Delta p)^k] = \mathcal{O}(\delta t)^2 \text{ for } k \geq 3. \tag{2}$$

As $N \to \infty$ we see that, over an infinitesimal timestep of length δt, Δp is approximately

normally distributed with

$$\Delta p \sim N(0, p(1-p)\delta t).$$

In other words, in the limit $\{p_t\}_{t\geq 0}$ follows the diffusion with generator

$$\mathcal{L}f(p) = \frac{1}{2}p(1-p)\frac{\partial^2 f}{\partial p^2},$$

or equivalently

$$dp_t = \sqrt{p_t(1-p_t)}dW_t \tag{3}$$

where $\{W_t\}_{t\geq 0}$ is a standard Brownian motion. Remember that in the passage to the limit we measured time in units of size N and so the appropriate approximation for a population of size N in *real* time units, denoted by τ, is

$$dp_\tau = \sqrt{\frac{1}{N}p_\tau(1-p_\tau)}dW_\tau$$

and backwards in time the coalescence of ancestral lineages is at rate $\frac{1}{N}\binom{k}{2}$. Crucially, *the application of the model requires knowledge of N.*

The stochastic force which drives the Wright-Fisher diffusion is called *genetic drift* and its strength is entirely determined by population size. There is considerable debate about the relative importance of genetic drift—which arises just from random *resampling* (reproduction) in a population—and other evolutionary forces, not least because a theory based on drift alone predicts a strong footprint of population size in data which is simply not present.

A basic observation is that genetic diversity is orders of magnitude lower than expected from census numbers and 'standard' genetic drift, Lewontin (1974), Gillepsie (2001). Something else is going on.

3. Genetic draft. So let's consider some of the other forces which might shape our genealogical trees. The first is *genetic structure*. In particular, we examine the influence of selection at one genetic locus on the genealogy of a sample from a neutral locus on the same chromosome. We shall suppose that the gene at the selected locus occurs in two forms which we label b, B. We start by finding the forwards in time analogue of equation (3)

for the frequency of b-alleles. As usual we start from a Wright-Fisher model.

Definition 3.1 (Wright-Fisher model with genic selection). A population of N genes occurring in two alleles (types) b and B evolves in discrete generations. Generation $(n+1)$ is formed from generation n by drawing N genes at random with replacement. If the *relative fitness* of the two alleles is $1+s:1$ and the frequency of b-alleles in the parental generation is p, then at each draw the probability that a b-allele is selected is

$$\frac{(1+s)p}{(1+s)p+(1-p)}.$$

Now we pass to what is known as the *weak selection* limit, obtained by assuming that s is $\mathcal{O}(\frac{1}{N})$. So writing $s = \frac{\sigma}{N}$ and using x_t to denote the proportion of b-alleles in the pop-

ulation, we perform the analogue of the calculation which led us to (1) to obtain this time

$$\begin{aligned}\mathbb{E}[\Delta x] &= \frac{(1+s)x}{(1+s)x+(1-x)} - x = \frac{sx(1-x)}{1+sx} \\ &= \frac{\sigma}{N}x(1-x) + \mathcal{O}(\frac{1}{N^2}) \\ &= \delta t \sigma x(1-x) + \mathcal{O}(\delta t^2)\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[(\Delta x)^2] &= \frac{1}{N^2} N \frac{(1+s)x}{(1+s)x+(1-x)}\left(1 - \frac{(1+s)x}{(1+s)x+(1-x)}\right) \\ &= \delta t x(1-x) + \mathcal{O}(\delta t^2).\end{aligned}$$

Once again we have

$$\mathbb{E}[(\Delta x)^k] = \mathcal{O}(\delta t)^2 \quad \text{for } k \geq 3.$$

As $N \to \infty$ we see that over an infinitesimal timestep of length δt,

$$\Delta x \sim N(\sigma x(1-x)\delta t, x(1-x)\delta t).$$

In other words (3) is replaced by

$$dx_t = \sigma x_t(1-x_t)dt + \sqrt{x_t(1-x_t)}dW_t.$$

In *real* time units we have

$$dx_\tau = sx_\tau(1-x_\tau)d\tau + \sqrt{\frac{1}{N}x_\tau(1-x_\tau)}dW_\tau. \tag{4}$$

We now use this equation to investigate the fate of a new selectively advantageous mutation arising in an otherwise neutral population. It will start at frequency $\frac{1}{N}$ in the population. Applying the theory of speed and scale for one-dimensional diffusions (see e.g. Karlin & Taylor 1981) to the diffusion (4), one finds that the probability that the new advantageous mutation will become *fixed* in the population, that is the probability that $\{x_\tau\}_{\tau \geq 0}$ hits one before it hits zero, given that it starts at $\frac{1}{N}$ is

$$P_{\text{fix}} = \frac{1-\exp(-2\sigma/N)}{1-\exp(-2\sigma)} \approx \frac{2s}{1-\exp(-2Ns)}. \tag{5}$$

Given that it does so, a Green function calculation which is given in detail in Etheridge et al. (2006), tells us that the expected value of the time T after its first appearance when it achieves fixation, in *real* time units, is

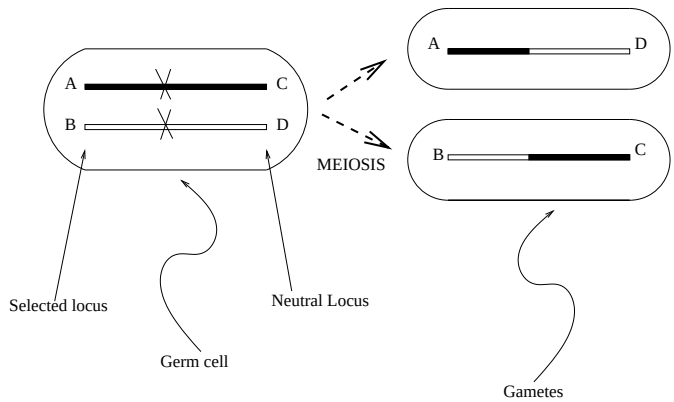
$$\mathbb{E}[T] = 2\frac{\log(Ns)}{s} + \mathcal{O}\left(\frac{1}{s}\right). \tag{6}$$

Moreover, from their Lemma 3.1, the variance $var[T]$ is $\mathcal{O}(1/s^2)$.

Definition 3.2 (Selective sweep). The process whereby the advantageous mutation increases to fixation is called a *selective sweep*.

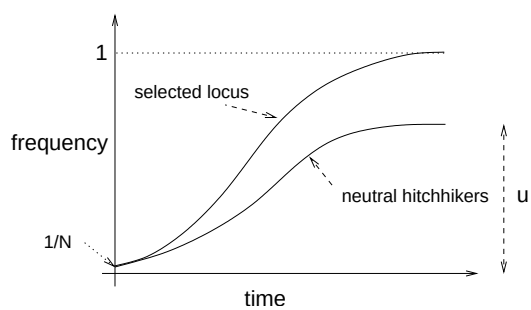
If chromosomes are passed down as indivisible blocks then during the course of the sweep the genetic type along the whole of the chromosome on which the mutation arises will also become fixed. But in a diploid population such as our own in which chromosomes are carried in *pairs* they are not passed down as indivisible blocks. So (ignoring Y-chromosomes) although you will have inherited one chromosome in the pair from your

mother and one from your father, it is not the case that the chromosome passed down by a parent to their offspring is an exact copy of one of the parental chromosomes but instead it is a mosaic of the two chromosomes in the parental pair. This is due to a process called *recombination*. Here, roughly, is how it works. During the reproductive process, each parent produces a large (effectively infinite) number of *germ cells*. These are cells all of the same genotype as the parent. These split into *gametes*. Gametes are cells that contain just one chromosome from each pair. It is the gametes that will fuse at random (one from the father, one from the mother) to produce the offspring. But during the process of splitting into gametes, which is called *meiosis*, the two chromosomes can exchange genetic material. This is illustrated in the cartoon below.



Such an event is called a *recombination* or *crossover* event. In general there could be more than one crossover event between two loci, indeed the number is often modelled as a Poisson process, but for our purposes we just need to know that there is some positive probability that the type at the neutral locus is inherited from a different parental chromosome from that at the selected locus.

As a result of the recombination events, during a selective sweep the frequency at a linked neutral locus may not be swept to fixation. Nonetheless, if the locus is sufficiently close to the selected locus that the probability of recombination is not too high, the frequency of the neutral allele which was fortunate enough to be associated with the selected allele when it arose will receive a boost from the selective sweep. This boost to the neutral locus is known as *genetic hitchhiking*, a term introduced by Maynard Smith & Haigh (1974). Schematically, the picture is something like the following:



We have used u to denote the proportion of neutral alleles at the end of the sweep descended from the individual associated with the favourable mutation when it first arose; that is u is the proportion of 'hitchhikers' in the population. Of course u is a random variable with a distribution depending on the strength of selection and the rate of recombination. Notice in particular that since we know from (6) that the sweep lasts $\mathcal{O}(\log(Ns)/s)$ generations, in order for the hitchhiking effect to be significant we require that the probability r of a recombination event is at most $\mathcal{O}(\frac{s}{\log(Ns)})$ in each generation. (This corresponds to the bound $r/s < 0.1$ given in Fay & Wu 2000.)

Gillespie (2000, 2001) investigated a model in which these strongly selected mutations which give rise to hitchhiking events occur at the points of a Poisson process. He assumes that selection is strong enough that the duration of the sweeps causing the hitchhiking events that affect a given locus is small compared to the time between them so that we can ignore the possibility that a locus will be subject to two simultaneous hitchhiking events. Then, in much the same way as we calculated the first two moments of the change of allele frequency over a single generation of the Wright-Fisher model, Gillespie calculated the first two moments in the change in allele frequency at the neutral locus over the course of a hitchhiking event. Again suppose that there are just two alleles at the neutral locus, a and A say. Suppose that the frequency of a-alleles before the sweep is p and consider the change Δp in frequency at the end of the sweep. Two things can happen:

1. If the a-allele hitchhikes, which happens with probability p, at the end of the sweep the frequency will be $u + p(1-u)$.
2. If the a-allele does not hitchhike, which happens with probability $1-p$, at the end of the sweep the frequency will be $(1-u)p$.

The expected change in frequency over the course of the sweep (the expectation is with respect to the u-random variable) is then

$$\mathbb{E}[\Delta p] = \mathbb{E}[p(u + p(1-u)) + (1-p)p(1-u)] - p = 0,$$

while

$$\mathbb{E}[(\Delta p)^2] = \mathbb{E}[p(u-pu)^2 + (1-p)(-pu)^2] = p(1-p)\mathbb{E}[u^2].$$

Notice that this looks a lot like genetic drift (equation 1) except that the rôle of $\frac{1}{N}$ is played by $\mathbb{E}[u^2]$—multiplied by the rate of sweeps to which the locus is subjected. This rate of sweeps is called the *substitution rate*.

Now Gillespie argues that $\mathbb{E}[u^2]$ times the substitution rate is much less sensitive to population size than $\frac{1}{N}$ and so this effect would allow us to carry over much of modern population genetics to a setting in which predictions are much less sensitive to population size than a theory based on drift alone.

Remark 3.3. At first sight the claim that substitution rate is rather insensitive to population size is counter-intuitive. If the chance of an advantageous mutation arising on an individual is μ in each generation, then the rate at which advantageous mutations arise across the whole population is $2N\mu$ and in view of equation (5), one is tempted to approximate the rate of substitution as $4N\mu s$ which obviously depends very strongly on population size. However, our calculation of the fixation probability for a new selectively

advantageous allele supposed that the mutation arose in an otherwise neutral population, but as the mutation rate increases we can expect to see *overlapping* sweeps. These will interfere with one another and in the absence of recombination we expect (see Cuthbertson et al. 2007) that if we assume what is known as multiplicative fitness, so the relative fitness of an individual carrying k mutations is $(1+s)^k$, then the rate of substitution grows at a rate proportional to $\log N$. In the hitchhiking world there *is* recombination, but we are only interested in a small portion of genome around the neutral locus in which recombination rates are low ($\mathcal{O}(\frac{1}{\log N})$) and so although the rate of adaptation will be increased by recombination, one does not expect it to grow significantly faster than $\mathcal{O}(\log N)$. Moreover, although we have modelled it as continuous, the genome is in reality discrete and so there is a bound on the number of distinct beneficial mutations that can arise in this small portion of genome around the neutral locus.

But Gillespie's *pseudo-hitchhiking model* as this is called differs in another important way from genetic drift. In the Wright-Fisher diffusion $\mathbb{E}[(\Delta p)^3] = \mathcal{O}(\delta t)^2$. Here that is *not* the case. In fact *all* moments of Δp will be of order δt. What this means for the genealogical trees relating individuals in a sample from the population is that whereas for the Kingman coalescent that corresponds to the Wright-Fisher diffusion we only ever saw two ancestral lines merging at a time, now there can be *multiple mergers* of ancestral lines.

In fact as we trace the ancestral lineages for a sample from the neutral locus back through the sweep, all those corresponding to genetic hitchhikers will merge into a single ancestor (corresponding to the founder of the sweep). The probability that any other lineages are involved in coalescence events during the course of the sweep is $\mathcal{O}(\frac{1}{\log N})$ and so to a first approximation we can approximate the genealogy at a neutral locus under the hitchhiking model by a coalescent in which at a coalescence event we see just a single merger. Conditional on u (which is independent at each successive event), for each ancestral lineage (independently) we flip a coin which comes up heads with probability u. All the ancestral lineages with a head merge into a single lineage. In particular *multiple* (by which we mean more than two) lineages can coalesce in a single event. This is a special case of something called a Λ-coalescent. Because in reality the sweep is not instantaneous, with probability of $\mathcal{O}(1/\log N)$ we see coalescence of lineages in the portion $1-u$ of the population whose type at the neutral locus does not derive from that of the originator of the sweep and so a better approximation for the genealogy is given by a coalescent which allows for *simultaneous (multiple) mergers*. Such coalescents are called Ξ-coalescents. A detailed discussion of the genealogy of the pseudo-hitchhiking model can be found in Durrett & Schweinsberg (2005). See also Durrett & Schweinsberg (2004), Schweinsberg & Durrett (2005) and Etheridge et al. (2006) for more on the effect of selective sweeps on genealogies.

4. Coalescents with multiple mergers. The processes known as Λ-coalescents were introduced independently by Pitman (1999) and Sagitov (1999). Like Kingman's coalescent they take their values among partitions of $\mathbb{N}$ and their laws can be prescribed by specifying the restriction to partitions of $[k] = \{1, 2, \ldots, k\}$ for each $k \in \mathbb{N}$.

DEFINITION 4.1 (Λ-coalescents). A Λ-*coalescent* $\{\pi(t)\}_{t\geq 0}$ is a Markov process taking its values among partitions of $\mathbb{N}$ with the property that for each k, the restriction to $[k]$, $\{\pi_k(t)\}_{t\geq 0}$, is also a Markov process and if there are currently n blocks in $\pi_k(t)$ then each transition involving j of the blocks merging into one happens at rate $\beta_{n,j}$ (which is *independent* of k) and these are the only possible transitions.

Generally $\pi(0)$ is taken to be the partition all of whose blocks are singletons. For our purposes the Λ-coalescent describes the ancestry of a population whose individuals are labelled by $\mathbb{N}$. Each block in the partition at time t corresponds to a single ancestor at time t before the present with the elements of the block being the descendants of that ancestor. Whereas for the Kingman coalescent the only transitions are mergers of pairs of blocks, for the Λ-coalescent there can be mergers of three or more lineages. The key point is that the conditions in Definition 4.1 ensure sampling consistency: the coalescent obtained by taking a sample of size k_1 and then sampling $k_2 < k_1$ of the individuals in the sample and looking at the restriction of the process to that subsample has the same distribution as the coalescent obtained by simply starting from a sample of size k_2.

If such a process is to exist, the parameters $\{\beta_{n,j}, 2 \leq j \leq n\}$ cannot be chosen arbitrarily.

THEOREM 4.2 (Pitman (1999)). *The* Λ*-coalescent of Definition 4.1 exists if and only if there is a finite measure* Λ *on* $[0,1]$ *for which*

$$\beta_{n,j} = \int_{[0,1]} u^{j-2}(1-u)^{n-j}\Lambda(du). \tag{7}$$

REMARK 4.3 (Recovering Kingman's coalescent). Notice that if $\Lambda = \delta_0$, the point mass on zero, then $\beta_{n,j}$ vanishes unless $j = 2$ (corresponding to two blocks, or ancestral lines, merging) in which case it is one so that we recover Kingman's coalescent.

More generally one can consider coalescents with *simultaneous* multiple collisions. Such coalescents were obtained as the limit of the ancestral processes of properly scaled population models (in much the same way as the Kingman coalescent arises as the rescaled ancestral process corresponding to the Wright-Fisher model) by Möhle & Sagitov (2001) who were able to classify all coalescents arising in this way as a limit of populations with exchangeable reproduction mechanisms. Schweinsberg (2000) independently obtained the same class of coalescents (without a passage to the limit) and characterised the possible rates of mergers of ancestral lineages in terms of a single measure Ξ on the infinite simplex

$$\Delta = \{(x_1, x_2, \ldots) : x_1 \geq x_2 \geq \ldots \geq 0, \sum_{i=1}^{\infty} x_i \leq 1\}.$$

The resulting coalescents are now known as Ξ-coalescents.

The Λ-coalescent specifies the genealogy of a sample from a forwards in time population model introduced by Bertoin & Le Gall (2003) that we shall call the Λ-Fleming-Viot process. This is closely related to work of Donnelly & Kurtz (1999). See also Birkner et al. (2005) and Berestycki et al. (2007) for explicit simultaneous constructions of the Λ-Fleming-Viot process and its genealogical trees.

The Λ-Fleming-Viot process takes its values among probability measures on $[0,1]$. We will describe it in terms of its generator, $\mathcal{R}$, acting on functions of the form

$$G(\rho) = \int f(x_1, \ldots, x_n)\rho(dx_n)\ldots\rho(dx_1), \tag{8}$$

where $n \in \mathbb{N}$ and $f : [0,1]^p \to \mathbb{R}$ is measurable and bounded. First we need some notation. If $x = (x_1, \ldots, x_n) \in [0,1]^n$ and $J \subseteq \{1, \ldots, n\}$ we write

$$x_i^J = x_{\min J} \text{ if } i \in J, \text{ and } x_i^J = x_i \text{ if } i \notin J, i = 1, \ldots, n. \tag{9}$$

Definition 4.4 (Λ-Fleming-Viot process). Let Λ be a finite measure on $[0,1]$. The *Λ-Fleming-Viot process* has generator

$$\mathcal{R}G(\rho) = \sum_{J\subseteq\{1,\ldots,n\},|J|\geq 2} \beta^{\Lambda}_{n,|J|} \int (f(x_1^J, \ldots, x_n^J) - f(x_1, \ldots, x_n))\rho(dx_n)\ldots\rho(dx_1), \tag{10}$$

where $\beta^{\Lambda}_{n,j}$ is defined in equation (7). When $\Lambda(\{0\}) = 0$ (so there is no Kingman component), this can also be written

$$\mathcal{R}G(\rho) = \int_{(0,1]} \int_{[0,1]} (G((1-u)\rho + u\delta_k) - G(\rho))\rho(dk)u^{-2}\Lambda(du). \tag{11}$$

The Λ-Fleming-Viot process is most easily understood in the case $\Lambda(\{0\}) = 0$. In that case one can think of it as follows. The population at time t is described by a probability measure $\rho(t)$ on the type space $[0,1]$. Take a Poisson point process on $\mathbb{R}_+ \times (0,1]$ with intensity measure $dt \otimes u^{-2}\Lambda(du)$ which picks times and sizes of jumps for our population process. At a jump time t with corresponding jump size u, a portion u of the population is killed and replaced by offspring of an individual chosen at random from $\rho(t-)$. Thus

$$\rho(t) = (1-u)\rho(t-) + u\delta_x$$

where x is chosen according to $\rho(t-)$. The duality with the Λ-coalescent is evident. To construct the genealogy of a sample from such a population, suppose that there are currently k ancestral lineages. We trace backwards in time until we first encounter a point of our Poisson process. Suppose that the jump size that this specifies is u. Then for each lineage, independently, we flip a coin which shows heads with probability u. All lineages with a head merge into a common ancestor and the process continues.

In Gillespie's pseudohitchhiking model, $dt \otimes \frac{1}{u^2}\Lambda(du)$ is determined by the rate of substitutions and the distribution of the hitchhiking effect encoded by u.

Evidently one can write down analogous population models whose genealogy is determined by the Ξ-coalescents, but we omit that here.

A population evolving according to the Λ-Fleming-Viot process forms a single mating unit, but real populations are spatially structured. For example they may be subdivided into discrete locations or distributed over a two-dimensional continuum. It is to spatial structure that we now turn our attention.

5. Spatial structure. The standard approach to space is to assume that the population is subdivided into *demes* (which one can think of as 'islands'), each of which is of

large constant size. Interaction between demes is through *migration* (or more accurately exchange) of individuals.

To get a feel for the effect that this will have on the genealogical trees for the population we first consider a very simple example. Consider a population that is divided into just two demes with migration between the two. This simple model arises as a model for a single population divided into two genetic types which are in approximate equilibrium in the population, but in which there is mutation between types. The Wright-Fisher model is adapted to this setting as follows:

DEFINITION 5.1 (Wright-Fisher model with migration). A population of size N is structured into two types, P and Q with $N_1 = N\omega_1$ of type P and $N_2 = N\omega_2$ of type Q. Each sub-population reproduces (independently) according to the neutral Wright-Fisher model except that now, after each reproduction step, a proportion of the population in each background is exchanged. In other words $\bar{\mu}_1 N_1$ individuals migrate from background one (P) to background two (Q) and $\bar{\mu}_2 N_2$ go the other way. In order to maintain constant population size in each background, we take $\bar{\mu}_1 N_1 = \bar{\mu}_2 N_2$.

Now let's establish the genealogy of a sample from such a population. First consider what happens to a single ancestral lineage as we trace backwards in time. Because the individuals in a given background are indistinguishable from one another, the probability that an individual of type P had a parent of type Q is just the proportion of individuals in the P-background after the migration step that had parents in the Q-background, namely $\frac{\bar{\mu}_2 N_2}{N_1} = \frac{\bar{\mu}_2 \omega_2}{\omega_1}$. Similarly, the probability that an individual in the Q background had a parent in the P background is $\frac{\bar{\mu}_1 \omega_1}{\omega_2}$. To obtain a diffusion limit we suppose that $\bar{\mu}_i = \frac{\mu_i}{N}$ where $N = N_1 + N_2$ is the total population size and we measure time in units of size N. Since the chance of a migration event and a coalescence event both affecting our ancestral lineages in a single generation is $\mathcal{O}(\frac{1}{N^2})$, in the diffusion timescale we only see coalescences between lineages in the same background. Our time unit is the *total population* size, as opposed to the population size in one of the backgrounds, so each pair of lineages currently in background i, coalesces at instantaneous rate $\frac{1}{\omega_i}$. We are implicitly assuming that $N\omega_i$ is *large* so that we never see multiple mergers. The genealogical trees for this model can then be described by a *structured (Kingman) coalescent*. As we trace backwards in time

1. ancestral lineages *migrate* from background one to background two at rate $\frac{\mu_2 \omega_2}{\omega_1}$ and from background two to background one at rate $\frac{\mu_1 \omega_1}{\omega_2}$,
2. any pair of lineages currently in background i *coalesces* at instantaneous rate $\frac{1}{\omega_i}$.

(For this and other simple extensions of Kingman's coalescent see Hudson 1990.) The key thing to note is that the rate of migration of ancestral lineages is weighted by the ratio of the population size in the two demes, so that backwards in time the migration mechanism is biased towards the more populous deme, and the rate of coalescence within a deme depends on population size there.

In more general spatial contexts, the standard approach is to generalise the simple two-deme model above to a collection of demes, indexed by some set I. The proportion of a-alleles in deme i at time t is denoted by $p_i(t)$ and is governed by Kimura's *stepping*

stone model (Kimura 1953):

$$dp_i = \sum_j m_{ij}(p_j - p_i)dt + \sqrt{\gamma p_i(1-p_i)}dW_i. \tag{12}$$

Here $m_{ij} = m_{ji}$ reflects migration between demes and the coefficient γ reflects the population size in each deme. The $\{W_i\}_{i\in I}$ are independent standard Brownian motions. In other words we have a system of interacting Wright-Fisher diffusions.

REMARK 5.2. We can more generally take γ to depend on i, reflecting different population sizes in different demes, but then since we are assuming that the population size in different demes is maintained we must also modify the migration mechanism to reflect the condition $\overline{\mu}_1 N_1 = \overline{\mu}_2 N_2$ in Definition 5.1.

The genealogical trees relating a sample from a population evolving according to Kimura's model are again determined by a structured coalescent which can most easily be described in terms of coalescing random walks. Ancestral lineages follow independent random walks (with rates determined by the coefficients $\{m_{ji}\}$) between the demes as we trace backwards in time but while in the same deme each pair of lineages coalesces at rate one. This result is essentially due to Shiga (1982), but the duality between coalescing random walks and the stepping stone model that he establishes is not as strong as we are claiming here. It is the more recent ideas of Donnelly & Kurtz (1999) that allow one to simultaneously construct the stepping stone model and its genealogical trees.

The stepping stone model is extremely popular, not least because it is characterised in terms of easily interpreted parameters and it is relatively easy to simulate the genealogies. However, for many biological populations it is unnatural to think of them as subdivided and for many years people have been looking for an analogue of the stepping stone model in a two-dimensional continuum.

In one spatial dimension one can apply the *diffusive rescaling* (so that the random walk governing migration of individuals converges to Brownian motion) to the stepping stone model to obtain a stochastic *partial* differential equation

$$dp = \frac{1}{2}\Delta p dt + \sqrt{\gamma p(1-p)}dW, \tag{13}$$

where W is now a space-time white noise (cf. Nagylaki 1978). But in two dimensions this equation has no solution; the white noise is 'too rough'. (See Walsh 1986 for an introduction to stochastic partial differential equations.) But anyway, in two dimensions, equation (13) is not what comes out of a diffusive rescaling. To see what *does* come out of the diffusive rescaling it is easiest to think about the structured coalescent. The coalescing random walks should converge to coalescing Brownian motions (and indeed in one dimension they do) but in two dimensions, two independent Brownian motions will never meet and so they never coalesce. We are just left with the Laplacian term (corresponding to migration in the population), the genetic drift disappears.

One approach to overcoming this difficulty is is to assume that the genealogical trees can be constructed from Brownian motions which coalesce at an instantaneous rate given by a function of their separation. The position of the common ancestor is generally taken

to be the midpoint between the two lineages immediately before the coalescence event (although other distributions are of course possible). Sadly this rather natural model has a number of deficiencies. For example, suppose that under this model one takes the genealogical tree relating a sample of size k and examines the subtree obtained by looking at a subsample of size two. The distribution of that subtree will *not* be the same as the distribution of the genealogical tree corresponding to a sample of size two. The reason is that whenever one of the two ancestral lineages is involved in a coalescence event in the full tree it will jump. Furthermore there is no corresponding *forwards* in time model for the evolution of the population.

One can of course argue that even a population evolving in a continuum can be approximated by a subdivided one provided the spatial subdivision is sufficiently fine, but this raises another important issue. A key assumption in the derivation of the stepping stone model from individual based models is that population size in each deme is large. However, in real biological populations the number of individuals living in a local neighbourhood will often be rather small. The structured coalescent, just as Kingman's coalescent, allows for only pairwise mergers of ancestral lineages, but if several individuals are sampled from one location and neighbourhood size is small then multiple coalescences (by which as usual we mean merging of at least three lineages) will become significant.

6. A new model. Very recently, Barton & Etheridge (2007) have introduced a new model for evolution in a spatial continuum. A significant advantage of the new model is that it allows us to explicitly incorporate certain large scale fluctuations in the population, but it also reflects finite local population density by allowing for multiple coalescences. Recall the basic observation that Gillespie was trying to address with the pseudohitchhiking model—genetic diversity is orders of magnitude lower than expected from census population size and genetic drift. One alternative explanation to the pseudo-hitchhiking effect is that real populations also experience *large scale* fluctuations in which the movement and reproductive success of many individuals are correlated. For example climate change has caused extreme extinction and recolonisation events that dominate the demographic history of humans and other species. It is plausible that such fluctuations are the cause of most of the observed drift, but we need adequate models before we can assess their importance relative to other possible causes.

For simplicitly we describe only a special version of our model which can be thought of as a spatial Λ-Fleming-Viot process with genealogical trees determined by a corresponding spatial Λ-coalescent. In this setting, after an extinction event a region is recolonised by the descendants of a single individual. More generally it would be natural to take a Poisson number of colonists and then the corresponding model would be a spatial Ξ-coalescent.

First we describe a prelimiting model (which would be an interesting object of study in its own right). Suppose a population is initially distributed as a Poisson point process with intensity λ say. Let $\mu(dr)$ be a measure (not necessarily finite, but not quite arbitrary as we shall see below) on $(0, \infty)$ and for each $r \in (0, \infty)$ let $\nu_r(du)$ be a probability measure

on $[0,1]$. The dynamics of our prelimiting model are described as follows:

1. Let Π be a Poisson Point Process on $\mathbb{R}_+ \times \mathbb{R}^2 \times \mathbb{R}_+$ with rate $dt \otimes dx \otimes \mu(dr)$.
2. If (t,x,r) is a point of Π, then at time t throw down a ball $B_r(x)$ of radius r and centre x in $\mathbb{R}^2$.
3. If the ball is empty do nothing. If not:
   (a) Choose an individual at random from those in $B_r(x)$ and select $u \in [0,1]$ at random according to ν_r.
   (b) For each individual in $B_r(x)$, independently flip a coin which shows heads with probability u and kill all those individuals with a head.
   (c) Throw down individuals with the same type as the selected individual (who may now be dead) according to an independent Poisson Point Process with intensity $u\lambda dx$ for $x \in B_r(x)$.

We shall refer to such events as reproduction events, but in some applications we have in mind they mimic rapid large-scale extinction-recolonisation events.

Of course any reproductive event has positive probability of leaving no individuals in a given region of space, but because neighbourhoods overlap, an empty neighbourhood can subsequently become recolonised and one can check (by a comparison with oriented percolation) that at least for sufficiently large λ the population will survive for all time. The difficulty is that it is not easy to write down explicitly the genealogical trees relating individuals in a sample from this population. An ancestor is necessarily in a non-empty patch of space and knowing that it is non-empty gives information about the rate at which it is hit by reproduction events as one traces back in time, but it is hard to find explicit expressions for this effect. We overcome this difficulty by considering a model in which the population density is infinite, but we retain some of the signature of a finite population by retaining the reproduction mechanism so that a non-trivial proportion of individuals in a neighbourhood are descended from a common ancestor. In particular, this will result in multiple coalescences of ancestral lineages.

Let us now describe this limiting model a little more precisely. We suppose that each individual in our population has a *type* taken from a set K (for example $K = [0,1]$) and a *location* in a space E. For illustration, here we take $E = \mathbb{R}^2$. The local population density is taken to be constant in space and time and we shall write $\rho(t,x,\cdot)$, or sometimes for brevity ρ_x, for the probability measure on K which describes the frequencies of different types among those individuals residing at the point x. The reproduction mechanism mirrors that for our discrete time model:

Definition 6.1 (Spatial Λ-Fleming-Viot process). The *spatial Λ-Fleming-Viot process* $\{\rho(t,x,\cdot), x \in \mathbb{R}^2, t \geq 0\}$ specifies a probability measure on the type space K for every $t \geq 0$ and every $x \in \mathbb{R}^2$. With the notation above, the dynamics of the process are as follows. At every point (t,x,r) of the Poisson point process Π we choose $u \in [0,1]$ independently according to the measure $\nu_r(du)$. We also select a point z at random from $B_r(x)$ and a type k at random according to $\rho(t-,z,\cdot)$. For all $y \in B_r(x)$,

$$\rho(t,y,\cdot) = (1-u)\rho(t-,y,\cdot) + u\delta_k.$$

Of course we must impose restrictions on the intensity measure if our process is to exist. To see what these should be, consider first the evolution of the probability measure $\rho(t,x,\cdot)$ defining the distribution of types at the point x. This measure experiences a jump of size $y \in A \subseteq [0,1]$ at rate

$$\int_{(0,\infty)} \int_A \pi r^2 \nu_r(du)\mu(dr).$$

By analogy with the Λ-Fleming-Viot process, we should like

$$\tilde{\Lambda}(du) = \int_{(0,\infty)} u^2 r^2 \nu_r(du)\mu(dr) \tag{14}$$

to define a finite measure on $[0,1]$. In fact, (weak) convergence of our individual based model to this limiting model as $\lambda \to \infty$ requires a bit more:

$$\tilde{\Lambda}(du) = \int_{(0,\infty)} u r^2 \nu_r(du)\mu(dr) \tag{15}$$

must define a finite measure on $[0,1]$.

Of course it is not enough to consider a single point. It has to be possible to 'fit together' the type distributions at different sites in a consistent way and the simplest way to guarantee that we can do this is to ensure the existence of a nice dual process describing, for each $n \in \mathbb{N}$, the distribution of lineages ancestral to a sample of size n from the population. Suppose then that a population evolves according to this model and consider the (backwards in time) dynamics of a *single* ancestral lineage. It evolves in a series of jumps with intensity

$$dt \otimes \int_{(|x|/2,\infty)} \int_{[0,1]} \frac{L_r(x)}{\pi r^2}\, u\, \nu_r(du)\mu(dr)dx$$

on $\mathbb{R}_+ \times \mathbb{R}^2$ where $L_r(x)$ is the area of $B_r(0) \cap B_r(x)$. If we want this to give a well-defined Lévy process, then we require

$$\int_{\mathbb{R}^2} (1 \wedge |x|^2) \Big(\int_{(|x|/2,\infty)} \int_{[0,1]} \frac{L_r(x)}{\pi r^2}\, u\, \nu_r(du)\mu(dr) \Big) dx < \infty. \tag{16}$$

Consider now lineages currently at separation $y \in \mathbb{R}^2$. They will coalesce if they are *both* involved in a replacement event which happens at instantaneous rate

$$\int_{(|y|/2,\infty)} L_r(y) \Big(\int_{[0,1]} u^2 \nu_r(du) \Big) \mu(dr). \tag{17}$$

Of course if two ancestral lineages do coalesce, then their common parent is located at a point selected at random from the ball involved in the reproduction event. Conceptually, this is readily extended to multiple lineages (where we will see multiple mergers). Notice that conditional on not having coalesced, the locations of ancestral lineages are *not* independent of one another. This is entirely analogous to the dependence between ancestral lineages in the coalescent for a continuous (finite) linear population suggested by Wilkins & Wakeley (2002) (see Wilkins 2004 for a two-dimensional analogue).

Only very preliminary calculations have been carried out for this model, but some elementary facts are known. For example, since population density is effectively constant, the path followed by a single lineage is the same forwards and backwards in time. An

elementary calculation also allows us to find the distribution of the time, $\tau(x)$, back to the most recent common ancestor of two individuals at separation x in the present day population. Fix $\theta > 0$ and write $\phi(x) = \mathbb{E}[\exp(-\theta\tau(x))]$, then

$$0 = (1-\phi(x)) \int_{[0,\infty)} \int_{[0,1]} u^2 L_r(x) \nu_r(du)\mu(dr) - \theta\phi(x)$$
$$+ \int_{\mathbb{R}^2} \int_{[0,\infty)} \int_{[0,1]} 2\left(u\frac{L_r(y)}{\pi r^2} - u^2 \frac{L_r(x,x+y)}{\pi r^2} \right) (\phi(x+y)-\phi(x))\nu_r(du)\mu(dr)dy,$$

where, as before, $L_r(x)$ is the area of the intersection of the ball of radius r centred on x and the ball of radius r centred on the origin and now $L_r(x, x+y)$ is the volume of the intersection of the three balls of radius r centred on the origin, x and $x+y$ respectively. Although unwieldy to deal with analytically, this quantity can be estimated numerically.

REMARK 6.2 (Generalisations). For concreteness we have presented only a simple form of the model. There are many generalisations.

1. Replace $\mathbb{R}^2$ by an arbitrary Polish space.
2. Choose a Poisson number of parents at each reproduction event instead of just one.
3. Choose the distribution of the parents at each reproduction event non-uniformly. For example, recolonisers after a large-scale extinction event may be more likely to come from the boundaries of the region.
4. Impose spatial motion of individuals not linked directly to the reproduction events.
5. Instead of replacing a portion u of individuals from a ball centred on x, replace individuals sampled according to some distribution (e.g. Gaussian) centred on x.
6. ... and many more.

In almost all existing models for spatial evolution, reproduction events affect the allele frequencies in one of a number of discrete neighbourhoods (demes). In our model, neighbourhoods are allowed to overlap. In fact, by replacing $\mathbb{R}^2$ by a discrete model and choosing the intensity of jumps appropriately, we recover the classical stepping stone and island models as special cases. Similarly, we can obtain the spatial version of the Λ-Fleming-Viot process and the corresponding spatial Λ-coalescent studied by Limic & Sturm (2006).

7. Demography. A major shortcoming of all the classical models which we have discussed so far is that they ignore the *demography* of the population. Recall that the key parameter in our theory of genetic drift was population size. This we took to be large and fixed. In our continuum stepping stone model too we have effectively taken local population density to be constant. Moreover, we have assumed that our population evolves in isolation, competing only with individuals of the same species. But real populations are not like this. The size of a population fluctuates both in time and space and different populations are in competition for the same resources.

It is widely believed that if one views populations over sufficiently large spatial and temporal scales, then there should be some averaging effect which would allow one to use classical population genetic models with constant population density but with *effective* parameters replacing the real population parameters. Various conditions are established

under which this really holds in Barton et al. (2002), but the usefulness of that result is limited due to a lack of explicit models for which the assumptions can be validated and the effective parameters calculated.

Natural populations interact with one another and with their environment in complex ways and we cannot hope to capture all of them in a single tractable mathematical model. Instead we try to isolate the effects of different aspects of these interactions through the study of simple 'toy' models. In this section we present several such models, all of which are compromises between fully spatial models and interacting particle systems. Many more details and some more general models can be found in Etheridge (2004) and Blath et al. (2007a,b).

In modelling a biological population it is natural to start from a branching process. However, in such models a finite population will either die out or grow without bound. In high dimensions, one can nonetheless obtain models in which population density is finite by allowing individuals to diffuse across an infinite domain, but in one and two dimensions (which is for many biological populations the most relevant), if one assumes that individuals diffuse either according to Brownian motion or a symmetric random walk then the spatial branching process predicts that if not extinct, at large times the population will develop clumps of arbitrarily large density and extent.

In a natural population one expects such clumping to be inhibited by limits on local resources. This effect can be introduced into our models by supposing that individuals living in crowded regions have lower reproductive success than those living in sparsely populated regions. Inspired by Bolker & Pacala (1997), the following class of models (and a continuum analogue) were introduced in Etheridge (2004):

$$dX_i(t) = \sum_j m_{ij}(X_j(t) - X_i(t))dt + \alpha(M - \sum_j \lambda_{ij} X_j(t))X_i(t)dt + \sqrt{\gamma X_i(t)}dB_i(t).$$

Here $X_i(t)$ denotes the size of the population in deme i at time t and $\{\{B_i(t)\}_{t\geq 0}, i \in \mathbb{Z}^d\}$ is a family of independent Brownian motions. The parameters $\{m_{ij}\}$ reflect migration of individuals just as in the stepping stone model. The noise is Feller's noise, corresponding to finite variance continuous state branching. The parameters $\{\lambda_{ij}\}$, which are positive, measure 'neighbourhood size'. It is shown in Etheridge (2004) that provided $m_{ij} > c\lambda_{ij}$ then for sufficiently large M the process will survive. This condition on the relative strengths of migration and competition was arrived at independently by Law et al. (2002). For the case when the competition is only within site, that is $\lambda_{ij} = 0$ for $i \neq j$, Hutzenthaler & Wakolbinger (2007) prove an ergodic theorem and also (under the same condition) show that there is a bound on the value of M below which the population will always go extinct.

This model is readily extended to competing species. Following Bolker & Pacala (1999) and Murrell & Law (2003), Blath et al. (2007a,b) assume the following strategies for survival for individuals in the population:

1. to colonise relatively unpopulated areas quickly,
2. to quickly exploit resources in those areas,
3. to tolerate local competition.

Suppose that there are just two different populations. Each can adopt a different combination of strategies for survival. We write $\{X(t)\}_{t\geq 0} = \{X_i(t), i \in \mathbb{Z}^d\}_{t\geq 0}$ and $\{Y(t)\}_{t\geq 0} = \{Y_i(t), i \in \mathbb{Z}^d\}_{t\geq 0}$ for our two populations, then in the model of Blath et al. (2007a,b) the populations evolve according to the following system of stochastic differential equations:

$$dX_i(t) = \sum_{j\in\mathbb{Z}^d} m_{ij}(X_j(t) - X_i(t))dt + \alpha\Big(M - \sum_{j\in\mathbb{Z}^d} \lambda_{ij}X_j(t) - \sum_{j\in\mathbb{Z}^d} \gamma_{ij}Y_j(t)\Big)X_i(t)dt + \sqrt{\sigma X_i(t)}dB_i(t), \qquad (18)$$

$$dY_i(t) = \sum_{j\in\mathbb{Z}^d} m'_{ij}(Y_j(t) - Y_i(t))dt + \alpha'\Big(M' - \sum_{j\in\mathbb{Z}^d} \lambda'_{ij}Y_j(t) - \sum_{j\in\mathbb{Z}^d} \gamma'_{ij}X_j(t)\Big)Y_i(t)dt + \sqrt{\sigma Y_i(t)}dB'_i(t), \qquad (19)$$

where $\{\{B_i(t)\}_{t\geq 0}, \{B'_i(t)\}_{t\geq 0}, i \in \mathbb{Z}^d\}$ is a family of independent Brownian motions. The (bounded non-negative) parameters m_{ij}, m'_{ij}, λ_{ij}, λ'_{ij}, γ_{ij}, and γ'_{ij} are all supposed to be functions of $\|i-j\|$ alone (where $\|\cdot\|$ denotes the lattice distance on $\mathbb{Z}^d$) and to vanish for $\|i-j\| > R$ for some $R < \infty$.

For the X-population, the first two strategies for survival listed above correspond to taking large m_{ij} and large αM, while the third corresponds to taking small λ_{ij} (conspecific competition) and γ_{ij} (interspecific competition). By varying M we can also model how efficiently the species uses the available resources: a species that can tolerate lower resource levels will have a higher value of M.

Although it is possible to show that there are parameter ranges for which there is long term coexistence of the two populations, the model is difficult to study and it is instructive to pass to a simpler one.

Suppose now that the neighbourhood over which each individual competes is just the site in which it lives so that the only interaction between different points in $\mathbb{Z}^d$ is through migration and assume that the competition parameters $\lambda_{ii}, \gamma_{ii}$ etc. are constant (and so we may omit the subscripts on them). In addition we suppose that the migration mechanism for the two populations is the same. Write

$$N_i(t) = X_i(t) + Y_i(t), \quad \text{and} \quad p_i(t) = \frac{X_i(t)}{N_i(t)}.$$

Then an application of Itô's formula (and some rearrangement) gives

$$dp_i(t) = \sum_j \frac{N_j(t)}{N_i(t)} m_{ij}(p_j(t) - p_i(t))dt + \alpha p_i(1-p_i)\big[M - \lambda N_i(t)p_i - \gamma N_i(t)(1-p_i)\big]dt - \alpha' p(1-p)\big[M' - \lambda' N_i(t)(1-p_i) - \gamma' N_i(t)p_i\big]dt + \sqrt{\frac{1}{N}p_i(t)(1-p_i(t))}\, dW_i(t),$$

where $\{\{W_i(t)\}_{t\geq 0}, i \in \mathbb{Z}^d\}$ is a family of independent Brownian motions. And conditioning on $N_i(t) \equiv N$ we obtain

$$dp_i(t) = \sum_j m_{ij}(p_j(t) - p_i(t))dt + sp_i(t)(1-p_i(t))(1-\mu p_i(t))dt + \sqrt{\frac{1}{N}p_i(t)(1-p_i(t))}\, dW_i(t), \qquad (20)$$

where

$$s = \alpha M - \alpha' M' + N(\alpha'\lambda' - \alpha\gamma)$$

and

$$\mu = \frac{(\alpha'\lambda' - \alpha\gamma)N + (\alpha\lambda - \alpha'\gamma')N}{\alpha M - \alpha' M' + (\alpha'\lambda' - \alpha\gamma)N}.$$

Equation (20) also has an interpretation from the genetics literature. There, $p_i(t)$ would be thought of as the number of individuals (chromosomes in this context) in the population of type a. These chromosomes are subject to selection, but not the genic selection that we described before which makes it advantageous to carry one of the two alleles, but *balancing selection*. There are some genes for which it is beneficial for an individual to carry one allele of each type (heterozygous advantage). To model this in our Wright-Fisher model with selection, we replace the constant parameter s in (4) by $s_1(1-p) - s_2 p$ to reflect the idea that with probability $1-p$ an a allele will be paired with an A individual in which case it confers the advantage s_1 and with probability p is paired with another a individual in which case it confers the *dis*advantage s_2. If $s_1, s_2 < 0$ then we obtain selection in favour of homozygosity.

If $\mu < 1$, then in each site i there is selection in favour of either the X-type or the Y-type according to whether $s > 0$ or $s < 0$. If $\mu > 1$, in each site i we have selection in favour of *heterozygosity* if $s > 0$ and selection in favour of *homozygosity* if $s < 0$. In the 'neutral' case, $s = 0$, we recover Kimura's stepping stone model.

In the special case when the two populations evolve symmetrically, equation (20) takes the simple form

$$dp_i(t) = \sum_j m_{ij}(p_j(t) - p_i(t))dt + sp_i(t)(1-p_i(t))(1-2p_i(t))dt + \sqrt{\frac{1}{N}p_i(t)(1-p_i(t))}\, dW_i(t). \quad (21)$$

For general s there is no convenient coalescent dual, but by first transforming the equations one can find an alternative duality with a system of *branching annihilating random walks*.

Definition 7.1 (Branching annihilating random walk). The Markov process $\{n_i(t), i \in \mathbb{Z}^d\}_{t\geq 0}$ with values $n_i(t) \in \mathbb{Z}_+$ and dynamics described by

$$\begin{array}{lll} \left\{ \begin{array}{l} n_i \mapsto n_i - 1, \\ n_j \mapsto n_j + 1 \end{array} \right. & \text{at rate } n_i m_{ij} & \text{(migration)} \\ n_i \mapsto n_i + m & \text{at rate } sn_i & \text{(branching)} \\ n_i \mapsto n_i - 2 & \text{at rate } \frac{1}{2}n_i(n_i-1) & \text{(annihilation)} \end{array}$$

is called a *branching annihilating random walk* with *offspring number* m and *branching rate* s.

The transformation that we require is simple. Let $x_i(t) = 1 - 2p_i(t)$. Then

$$dx_i(t) = \sum_j m_{ij}(x_j(t) - x_i(t))dt + \frac{1}{2}s(x_i^3(t) - x_i(t))dt - \sqrt{(1-x_i^2(t))}dW_i(t). \quad (22)$$

LEMMA 7.2. *The system* (22) *is dual to branching annihilating random walk with branching rate* $s/2$ *and offspring number two, denoted* $\{n_i(t), i \in \mathbb{Z}^d\}_{t\geq 0}$, *through the duality relationship*

$$\mathbb{E}[\underline{x}(t)^{\underline{n}(0)}] = \mathbb{E}[\underline{x}(0)^{\underline{n}(t)}],$$

where

$$\underline{x}^{\underline{n}} \equiv \prod_{i\in\mathbb{Z}^d} x_i^{n_i}.$$

Cardy and Täuber (1996), (1998) consider the branching annihilating random walk model of Definition 7.1. Their results are not rigorous, but if true, the implications for the system (21) are as follows. For $\mu = 2$ and $d = 1$ there is a critical value $s_0 > 0$ such that the populations will both persist for all time with positive probability if and only if $s > s_0$. In $d = 2$, there is positive probability that both populations will persist for all time if and only if $s > 0$. For $d \geq 3$ this probability is positive if and only if $s \geq 0$. Roughly speaking, for $d \geq 2$, if there is a homozygous advantage, then the population will initially form homogenic clusters, but ultimately it will be the interactions at the cluster boundaries that dominate and one type will go extinct. In the heterozygous advantage case, there will be long term coexistence of species. In one dimension, the heterozygous advantage must be 'sufficiently strong' if we are to see coexistence.

One expects the same result to be true when $\mu \neq 2$. Notice then that we have selection in favour of heterozygosity precisely when

$$(\alpha\lambda_{ii} - \alpha'\gamma'_{ii})N > \alpha M - \alpha' M', \quad \text{and} \quad (\alpha'\lambda'_{ii} - \alpha\gamma_{ii})N > \alpha' M' - \alpha M.$$

Comparing the quantities $\alpha'\lambda'_{ii} - \alpha\gamma_{ii}$ and $\alpha\lambda_{ii} - \alpha'\gamma'_{ii}$ tells us about the relative effectiveness of the X and Y populations as competitors. If the first is smaller, then the X-population is a less effective competitor. However, provided that $\alpha M > \alpha' M'$, we can even allow it to be negative and we expect to have positive probability of survival for the X-population. This reflects a competition-colonisation tradeoff.

8. Where next? The primary motivation underlying all the work described above is that a theory of evolution based on genetic drift alone predicts dramatically more genetic variability than we observe in data. If we are to correctly interpret that data then we must find ways to distinguish the different causes of reduced genetic diversity.

The new model described in §6 presents a promising platform in which to do this, at least for some evolutionary scenarios, but so far it has barely been explored. The first major issue to resolve is that we must establish the class of individual based models that can be approximated in this unifying framework. Second we must find statistics of the model that distinguish for example the effects of large scale extinction recolonisation events from those of natural selection. What is the relative importance of the (rare) large scale events and the (frequent) small scale events and can we find a footprint of both effects in data?

Demographic models too are a rich source of open problems. The key issue is to find a mathematically tractable model which nonetheless retains sufficient flexibility to incorporate the main strategies for survival that a species can adopt. Even our toy models

have proved surprisingly resistant to mathematical analysis. For example, in spite of considerable efforts, although through studying the system (20) directly we have been able to show that for *sufficiently large s* there is positive probability that branching annihilating random walk with offspring number two will survive for all time, we have *not* been able to reproduce the full results of Cardy and Täuber.

Finally let us mention that very few genetic models incorporate demography in a sophisticated way. For example, one might use a model for population size expansion to model a time-varying coalescence rate in Kingman's coalescent, but rather few spatial models incorporate fluctuations in the local population size. However, as explained in detail in Etheridge (2006), even small fluctuations in local population size can have a large effect on the genealogical trees relating individuals in a sample from such a population. So although ecological and evolutionary models are traditionally treated separately, one of the most important challenges that faces us is to find ways to combine demography and genetics into a single tractable model.

Acknowledgements. The author would like to thank the anonymous referee for a detailed and helpful report which led to many clarifications.

References

[1] N. H. Barton, F. Depaulis, and A. M Etheridge, *Neutral evolution in spatially continuous populations*, Theor. Pop. Biol. 61 (2002), 31–48.

[2] N. H. Barton and A. M. Etheridge, *A continuum stepping stone model*, manuscript, 2007.

[3] J. Berestycki, N. Berestycki, and V. Limic, *On small time asymptotics of* Λ*-coalescents*, preprint, 2007.

[4] J. Bertoin and J.-F. Le Gall, *Stochastic flows associated to a coalescent process*, Prob. Theor. Rel. Fields 126 (2003) 261–288.

[5] M. Birkner, J. Blath, M. Capaldo, A. M. Etheridge, M. M.öhle, J. Schweinsberg, and A. Wakolbinger, *Alpha-stable branching and Beta-coalescents*, Elect. J. Probab. 10 (2005), 303–325.

[6] J. Blath, A. M Etheridge, and M. E Meredith, *Coexistence in locally regulated competing populations and survival of branching annihilating random walk*, Ann. Appl. Probab. 17 (2007), 1474–1507.

[7] J. Blath, A. M Meredith, and M. E Meredith, *Coexistence in locally regulated competing populations and survival of branching annihilating random walk (full version)*, Technical University of Berlin Preprint, 2007.

[8] B. M. Bolker and S. W. Pacala, *Using moment equations to understand stochastically driven spatial pattern formation in ecological systems*, Theor. Pop. Biol. 52 (1997), 179–197.

[9] B. M. Bolker and S. W. Pacala, *Spatial moment equations for plant competition: Understanding spatial strategies and the advantages of short dispersal*, American Naturalist 153 (1999), 575–602.

[10] J. L. Cardy and U. C. Täuber, *Theory of branching and annihilating random walks*, Phys. Rev. Lett. 77 (1996), 4780–4783.

[11] J. L. Cardy and U. C. Täuber, *Field theory of branching and annihilating random walks*, J. Stat. Phys. 90 (1998), 1–56.

[12] C. Cuthbertson, A. M. Etheridge, and F. Yu, *Asymptotic behaviour of the rate of adaptation*, preprint, 2007.

[13] P. J. Donnelly and T. G. Kurtz, *Particle representations for measure-valued population models*, Ann. Prob. 27 (1999), 166–205.

[14] R. Durrett and J. Schweinsberg, *Approximating Selective Sweeps*, Theoretical Population Biology 66 (2004), 129–138.

[15] R. Durrett and J. Schweinsberg, *A coalescent model for the effect of advantageous mutations on the genealogy of a population*, Stoch. Proc. Appl. 115 (2005), 1628–1657.

[16] A. M. Etheridge, *Survival and extinction in a locally regulated population*, Ann. Appl. Probab. 14 (2004), 188–214.

[17] A. M. Etheridge, *Evolution in fluctuating populations*, in: A. Bovier, F. Dunlop, A. van Enter, F. den Hollander, and J. Dalibard (eds.), Mathematical Statistical Physics, Lecture notes of the Les Houches Summer School 2005, Elsevier, 2006.

[18] A. M. Etheridge, P. Pfaffelhuber, and A. Wakolbinger, *An approximate sampling formula under genetic hitchhiking*, Ann. Appl. Probab. 16 (2007), 685–729.

[19] J. C. Fay and C.-I. Wu, *Hitchhiking under positive Darwinian selection*, Genetics 155 (2000), 1405–1413.

[20] J. Gillespie, *Genetic drift in an infinite population: the pseudohitchhiking model*, Genetics 155 (2000), 909–919.

[21] J. Gillespie, *Is the population size of a species relavant to its evolution?*, Evolution 55 (2001), 2161–2169.

[22] R. Hudson, *Gene genealogies and the coalescent process*, Oxford Surveys in Evolutionary Biology 7 (1990), 1–44.

[23] M. Hutzenthaler and A. Wakolbinger, *Ergodic behaviour of locally regulated populations*, Ann. Appl. Probab. 17 (2007), 474–501.

[24] S. Karlin and H. M. Taylor, *A Second Course on Stochastic Processes*, Academic Press, 1981.

[25] M. Kimura, *Stepping stone model of population*, Ann. Rep. Nat. Inst. Genetics Japan 3 (1953), 62–63.

[26] J. F. C. Kingman, *The coalescent*, Stoch. Proc. Appl. 13 (1982), 235–248.

[27] R. Law, D. J. Murrell, and U. Dieckmann, *On population growth in space and time: spatial logistic equations*, Ecology, 2002.

[28] R. C. Lewontin, *The Genetic Basis of Evolutionary Change*, Columbia University Press, New York, 1974.

[29] T. Shiga, *Continuous time multiallelic stepping stone models in population genetics*, J. Math. Kyoto Univ. 22 (1982), 1–40.

[30] V. Limic and A. Sturm, *The spatial Lambda-coalescent*, Electron. J. Probab. 11 (2006), 363–393.

[31] M. Möhle and S. Sagitov, *A classification of coalescent processes for haploid exchangeable models*, Ann. Probab. 29 (2001), 1547–1562.

[32] D. J. Murrell and R. Law, *Heteromyopia and the spatial coexistence of similar competitors*, Ecology Letters 6 (2003), 48–59.

[33] T. Nagylaki, *A diffusion model for geographically structured populations*, J. Math. Biol. 6 (1978), 375–382.

[34] J. Pitman, *Coalescents with multiple collisions*, Ann. Probab. 27 (1999), 1870–1902.

[35] S. Sagitov, *The general coalescent with asynchronous mergers of ancestral lines*, J. Appl. Probab. 26 (1999), 1116–1125.

[36] J. Schweinsberg, *Coalescents with simultaneous multiple collisions*, Electron. J. Probab. 5 (2000), 1–50.

[37] J. Schweinsberg and R. Durrett, *Random partitions approximating the coalescence of lineages during a selective sweep*, Ann. Appl. Probab. 15 (2005), 1591–1651.

[38] J. Maynard Smith and J. Haigh, *The hitch-hiking effect of a favourable allele*, Genet. Res. 23 (1974), 23–35.

[39] J. B. Walsh, *An introduction to stochastic partial differential equations*, in: École d'été de probabilités de Saint Flour, Lecture Notes in Mathematics 1180, 1986, 265–439.

[40] J. F. Wilkins, *A separation of timescales approach to the coalescent in a continuous population*, Genetics 168 (2004), 2227–2244.

[41] J. F. Wilkins and J. Wakeley, *The coalescent in a continuous, finite, linear population*, Genetics 161 (2002), 873–888.

STOCHASTIC MODELS IN BIOLOGICAL SCIENCES
BANACH CENTER PUBLICATIONS, VOLUME 80
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WARSZAWA 2008

LINKING POPULATION GENETICS TO PHYLOGENETICS

PAUL G. HIGGS

Department of Physics and Astronomy, McMaster University
Hamilton, Ontario L8S 4M1, Canada
E-mail: higgsp@mcmaster.ca

Abstract. Population geneticists study the variability of gene sequences within a species, whereas phylogeneticists compare gene sequences between species and usually have only one representative sequence per species. Stochastic models in population genetics are used to determine probability distributions for gene frequencies and to predict the probability that a new mutation will become fixed in a population. Stochastic models in phylogenetics describe the substitution process in the single sequence that represents a species. These latter models are of great practical use in inferring evolutionary trees, but they ignore population genetics. Here, we consider the way that the two fields are linked. In a finite population with a low mutation rate, most sites in a gene are monomorphic; therefore it is a reasonable approximation to represent a species by a single sequence and to ignore within population variability. We show that population genetics theory can be used to predict the substitution rates that will be seen in phylogenetic models. However, a substitution in the phylogenetic sense is really the fixation of a new mutation in the population genetics sense. Substitution rates are dependent on mutation, selection and random drift, not just on mutation. An appreciation of the relationship with population genetics is important when designing new empirical models for phylogenetics, for example, paired-site models for RNA sequences with secondary structure and codon-based models for protein-coding regions of DNA.

1. Introduction. This paper arises from the school on "Stochastic Models in Biological Sciences" held at the Banach Centre in 2006. It is intended as a pedagogical introduction to the population genetics models of allele frequencies and the phylogenetics models of substitution rates. The aim is to show that these two types of models can be brought together, and that an understanding of population genetics is useful when developing new models in phylogenetics.

2000 *Mathematics Subject Classification*: 92D15.

Key words and phrases: population genetics, phylogenetics.

This research was supported by the Canada Research Chairs organization and by the Natural Sciences and Engineering Research Council of Canada.

The paper is in final form and no version of it will be published elsewhere.

 © Instytut Matematyczny PAN, 2008

Population genetics is the study of the variability of genes within populations. This is a well-established discipline with a strong theoretical foundation dating back to the first half of the twentieth century. A classic problem in population genetics is the calculation of the frequency distribution of alleles at a genetic locus. (A locus is a site on a chromosome where a particular type of gene is found and an allele is one of a number of alternative versions of the gene that can be present at that locus.) By taking a sample of individuals from a population, it is possible to estimate the frequency of each of the alternative alleles in the population. The frequency of an allele will vary in time due to mutation, selection and random drift. Stochastic models are used to study the way allele frequencies change, and to predict the distribution of frequencies that is to be expected in different circumstances.

Molecular phylogenetics is the use of nucleic acid or protein sequence data from different species to construct evolutionary trees. Typically, a single sequence from each species is used. The variability within each species is neglected, because differences in gene sequences between individuals in a species are assumed to be smaller than differences between species. Stochastic models are also central to molecular phylogenetics. Phylogenetic models of DNA sequence evolution assume that substitutions of one nucleotide by another occur as random events at points in a gene sequence. The sequences of different species diverge over time due to the accumulation of many of these random substitutions independently in different lineages. The models are defined in terms of a substitution rate matrix in which the rates of different types of substitution depend on a number of parameters whose values can be estimated by fitting the model to sequence data.

Population genetics was founded before the era of molecular biology. It is possible to discuss the segregation of alleles for round and wrinkled peas or brown and blue eyes in a meaningful way without even knowing that these genes are encoded by DNA. The advent of experimental techniques for detecting alternative alleles, such as allozymes and restriction fragment length polymorphisms, led to much more detailed molecular input to population genetics, but these techniques still did not give the full sequence. More recently, it has become possible to sequence DNA from many individuals in a population, and to detect the specific sites in a gene where variability exists (known as single nucleotide polymorphisms - SNPs). This type of information is useful for medical studies, and in some cases, diseases can be linked to specific mutations. Because of the medical relevance, intraspecific sequence data is now becoming more common for humans. In most other species, we are still limited to only one sequence of any given gene, and we must presume that this is a typical representative of the species. Nevertheless, variability at the sequence level must exist in other species, as it does in humans.

The models of population genetics are defined from a mathematic point of view and are as close as one comes in biology to first-principles theory. On the other hand, the models of phylogenetics are empirical and have been built up to describe phenomena seen in real sequence data. The important results regarding the probabilities of fixation of new alleles as a function of mutation, selection and drift have been understood for a long time. However, until recently, these things have largely been ignored in practical phylogenetic studies. Substitution rate parameters in phylogenetic models are usually estimated merely

with the goal of obtaining the phylogeny, and little attempt has been made to interpret the substitution rates in terms of population genetics. Currently, with the increasing availability of both inter- and intra-specific sequence data, the two fields are beginning to touch on one another more closely than was formerly the case. Phylogenetic models are becoming increasingly more complex, and this allows a closer link to population genetics in the way the models are defined. Section 2 of this paper gives a brief introduction to the ideas of gene frequency distributions and fixation probabilities that are key aspects of population genetics. In Sections 3, 4 and 5, simple models of DNA sequence evolution from phylogenetics are introduced, and simple simulations are presented as illustrations of the way the two types of model are related. Sections 6 and 7 discuss more recent models that deal with sequences coding for RNAs and proteins. We consider the way that population genetics issues, like selection, fixation, and compensatory mutations, are relevant to evolution of these molecules and the way that these effects can be built into more realistic models of sequence evolution for use in molecular phylogenetics.

2. Stationary gene frequency distributions. We will consider a standard model in population genetics for the gene frequency distribution at a locus with four alleles. This applies to a single site in a DNA sequence that can be either A, C, G or T, under the assumption that the site is neutral (i.e. there is no difference in fitness between the nucleotides). Let $n_i(t)$ be the number of gene copies in the population at generation t that have base i at that site. We will follow the convention that the total number of copies of the gene is $2N$. This applies to a diploid splecies where N is the population size, or to a haploid species with population $2N$. The frequencies of the bases are $x_i(t) = n_i(t)/2N$. The Wright-Fisher model deals with constant population size and non-overlapping generations. Each gene copy at time $t+1$ is descended from one randomly chosen copy at time t. Let there be a probability u that there is a mutation from any base to any other base in one generation. The probability, a_i, that a new gene copy has i at this site is the probability that its parent was i and there was no mutation plus the probability that its parent was something other than i and there was a mutation to i.

$$a_i = x_i(t)(1-3u) + \sum_{j \neq i} x_j(t)u \tag{1}$$

Each gene copy is sampled independently by copying from the previous generation. Therefore, the probability distribution for the numbers of copies of each allele at the next generation (time $t+1$) is a multinomial distribution:

$$p(n_A, n_C, n_G, n_T) = \frac{(2N)!}{n_A! n_C! n_G! n_T!} a_A^{n_A} a_C^{n_C} a_G^{n_G} a_T^{n_T}. \tag{2}$$

For clarity, we have omitted the $t+1$ in all the $n_i(t+1)$ in equation 2. Using equations 1 and 2 repeatedly allows us to simulate a population history over many generations. If we keep track of the value of $n_i(t)$ at each time for any one of the four bases, we can calculate the probability, $\Phi(n)$, that $n_i(t) = n$. As $2N \gg 1$, this can be viewed as a continuous probability distribution $\phi(x)$, where $x = n/2N$. After a long time, $\phi(x)$ converges to the stationary distribution in equation 3, which can be calculated using diffusion theory (see

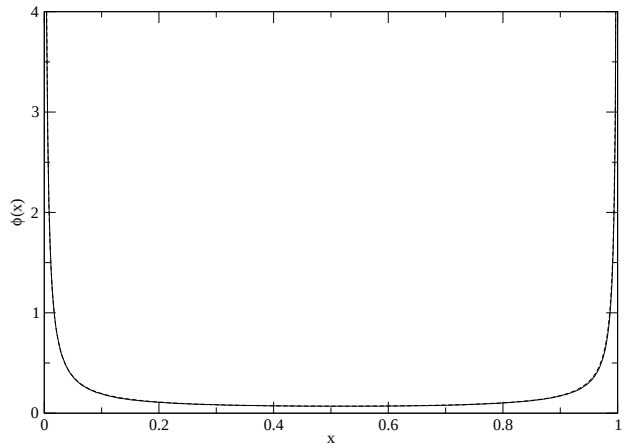


Fig. 1. Stationary frequency distribution for the case $\theta = 0.025$. The solid line is calculated from equation 3. The result from simulation of the the model for 10^8 generations with $2N = 500$ and $u = 2.5 \times 10^{-5}$ is superimposed as a dashed line, and is almost indistinguishable.

Wright [1], chapter 13, Crow and Kimura [2], chapter 9, or Durrett [3], chapter 1).

$$\phi(x) = \frac{\Gamma(4\theta)}{\Gamma(\theta)\Gamma(3\theta)} x^{\theta-1}(1-x)^{3\theta-1} \tag{3}$$

The parameter $\theta = 4Nu$ controls the shape of this distribution. The mutation rate, u, is very small in real cases (often quoted as of order 10^{-9} per site per generation), whereas the effective population size might be only around 10^4. Usually, therefore, $\theta \ll 1$, and both x and $(1-x)$ are raised to negative powers in equation 3. Hence, $\phi(x)$ has a U-shaped distribution that diverges as x tends to 1 or 0. The gamma functions in equation 3 are just constants that ensure the probability distribution is normalized. Figure 1 shows an example of the shape of the distribution for $\theta = 0.025$.

When $\theta \ll 1$, most of the weight of the probability distribution is close to 0 or to 1. As the four bases are equivalent in this model, there is a probability of 1/4 that any base will have frequency close to 1, and a probability of 3/4 that it will have frequency close to 0. A site is described as monomorphic if almost all the population has the same base at this site. For concreteness, we will say it is monomorphic if the most common base has a frequency greater than 95%. The site is polymorphic if its most common base has frequency less than 95%. If θ is small, most sites will be monomorphic most of the time, but they will occasionally pass through periods of polymorphism as they flip from being monomorphic for one base to being monomorphic for some other base.

The process of flipping from one state to another is called fixation. Consider a site that is completely monomorphic - let's say 100% A. If a mutation to a C occurs, there is just one gene copy that has C at this site. The frequency of C in the population will vary in time due to random drift (i.e. the multinomial sampling described by equation 2). The C allele has two possible fates. With a very small probability, p_{fix}, it will spread through the population until x_C is close to 1. In this case, we say that the C mutation

has become fixed in the population. However, with a probability $1 - p_{fix}$, random drift leads to the disappearance of C, and the population remains dominated by the A allele.

The probability of fixation of a neutral mutation, given that it begins with just one copy out of $2N$, can be shown to be $p_{fix} = 1/2N$ (Crow and Kimura [2] chapter 8). This can be calculated from diffusion theory, but it can also be obtained with the following simple argument. If we trace the lines of any two gene copies in the current population backwards in time, there comes a point where they coalesce, i.e. the two gene copies have the same ancestor. If we trace back the lines of descent of all genes in the current generation, they eventually coalesce to the a single gene copy that is the last common ancestor (LCA) of the whole of the current population. Now suppose a mutation occurred in the LCA generation. If it occurred in the gene copy that is the LCA itself, then it becomes fixed in the population at the current generation. If it occurs in any of the $2N - 1$ gene copies that is not the LCA, then it dies out before the current generation. Hence $p_{fix} = 1/2N$.

We can now calculate the probability that the population flips between the four alternative monomorphic states. For the population in the A state, the mean number of C mutations per generation in the whole population is $2Nu$. Each of these has a probability $1/2N$ of being fixed. Thus, the rate of flipping from state A to state C (or between any other two bases) is $r = 2Nu \times 1/2N = u$. This is a fundamental property of neutral evolution: the rate of evolution of the population is equal to the rate at which mutations occur in a single individual. It should be borne in mind that this is only true for a neutral mutation. For an advantageous mutation, $r > u$, and for a deleterious mutation, $r < u$.

3. Can the population be represented by a single sequence? We now begin to see the potential link between the population genetics and phylogenetics viewpoint. In phylogenetics, it is assumed that a population can be represented by a single gene sequence and that variation within the population is not important. If most sites in a gene are monomorphic, then this assumption seems reasonable at first sight. A single DNA site can be in states i = A, C, G or T. A phylogenetic model describing the evolution of the site is defined by a rate matrix r_{ij} in which each off-diagonal element is the rate of substitution from state i to state j. The diagonal elements are defined as $r_{ii} = -\sum_{j \neq i} r_{ij}$. It is necessary to insist on the distinction between substitutions and mutations. In a population genetics model, a mutation is a change in a single gene copy (due to replication error or chemical damage) that can be passed on to the next generation. In a phylogenetics model, a substitution is a change in the sequence that is being used to represent the population. A substitution arises as the result of the fixation of a mutation. Substitution rates are influenced by random drift and selection, as well as mutation.

The fundamental quantities to be calculated in a phylogenetic model are the transition probabilities $P_{ij}(t)$, defined as the probability that the site will be in state j after time t given that it was initially in state i. These probabilities satisfy the rate equation:

$$\frac{dP_{ij}}{dt} = \sum_k P_{ik} r_{kj}. \tag{4}$$

For simple choices of the r_{ij} matrix, this equation can be solved analytically. For more general cases, it can be solved numerically.

The simplest model of this type is the Jukes-Cantor (JC) model [4], in which $r_{ij} = u$, for all $i \neq j$ and $r_{ii} = -3u$. In this section, u is a substitution rate, i.e. it is the rate at which the sequence representing the population changes. In the previous section, u was a mutation rate. However, we saw that fixation and mutation rates are equal for neutral mutations; therefore, it should be possible to link the two models directly and have u be the same in both models. We now show that this *is* possible, but it is necessary to be careful.

If we simulate the evolution of a population evolving according to the Wright-Fisher model (as in section 2), we can use it to generate a time series, such as GGGGGTT-TAAAAATTTTCCCCCCAA..., which lists the state the population is in at each generation. We will consider three different ways of choosing the base to represent the state of the population below, but the application of the JC model to the time series is the same in each case. The JC model is a Markov model where it is assumed that the probability of being in a given state at any point in time depends only on the state at the previous generation. The transition probabilities in a fixed period of 1 generation are $P_{ij}(1)$, obtained from solution of equation 4. The time series can be summarized by the matrix of observed numbers of substitutions S_{ij}, which is simply the count of the number of times that state j follows state i in the time series. It is straightforward to calculate $P_{ij}(t)$ by solving equation 4 for the JC model (detailed solutions are given by Li [5], Durrett [3] and Higgs and Attwood [6]). The probability that the base at time t is the same as the initial base is $P_{ii}(t)$, and the probability that it is different from the initial base, irrespective of the which two bases these are, is $D(t)$, where:

$$D(t) = 1 - P_{ii}(t) = \frac{3}{4}(1 - \exp(-4ut)). \tag{5}$$

The total number of steps in the time series is $S_{tot} = \sum_i \sum_j S_{ij}$. The number of times that i and j are different is $S_{dif} = \sum_i \sum_{j \neq i} S_{ij}$. Therefore, the observed probability that the states are different at two successive steps is $\hat{D} = S_{dif}/S_{tot}$. The 'hat' denotes the value estimated from the data. By substituting $\hat{D}$ into equation 5 and inverting it, we obtain an estimate of ut.

$$\widehat{ut} = -\frac{1}{4}\ln(1 - 4\hat{D}/3). \tag{6}$$

It is not possible to estimate u and t separately from the observed transition matrix because transition probabilities are dimensionless, and they always depend on products of rates and times.

We simulated the Wright-Fisher model for 10^8 generations with parameters as in Figure 1 and used this to generate time series, from which we estimated ut by the method above. The most obvious way to generate the time series is to select one random individual from the population at each generation and to use the nucleotide from this individual to represent the population. This corresponds to the usual case in phylogenetics, where only one sequence from a species is available. A second method is to use the consensus state of the population, i.e. the nucleotide that is most frequent in the population at

Table 1. Alternative methods of estimating the substitution rate u from simulation data

method	interval t	S_{tot}	S_{dif}	$\hat{D}$	ut	u
random individual	1	10^8	6803246	0.0680	2.38×10^{-2}	2.38×10^{-2}
consensus state	1	10^8	254002	2.54×10^{-3}	8.48×10^{-4}	8.48×10^{-4}
counting fixations	1	10^8	7384	7.38×10^{-5}	2.46×10^{-5}	2.46×10^{-5}
random individual	10^3	10^5	13311	0.1331	0.0488	4.88×10^{-5}
consensus state	10^3	10^5	9416	0.0942	0.0335	3.35×10^{-5}
counting fixations	10^3	10^5	7154	0.0715	0.0251	2.51×10^{-5}
random individual	10^4	10^4	4952	0.4952	0.270	2.70×10^{-5}
consensus state	10^4	10^4	4772	0.4772	0.253	2.53×10^{-5}
counting fixations	10^4	10^4	4705	0.4705	0.247	2.47×10^{-5}

each generation. A third method is by counting the fixations, which can be done in the following way. If a population is monomorphic for base i (i.e. $x_i > 0.95$), we use i as the base in the time series. We then continue to use i in the time series until such point as the population becomes monomorphic for another base j. In this method, the population is still considered to be in state i in the period during which it is polymorphic, and the transition to state j is only deemed to have occurred when the population becomes monomorphic for j.

These three methods generate different time series from the same simulation. The estimates of ut from the three time series are shown in the top three lines of Table 1. We know in the simulation that the samples were taken at an interval of $t = 1$ generation. Thus $ut = u$ in the table. The true mutation rate is $u = 2.5 \times 10^{-5}$. The result obtained from counting the fixations is consistent with this, as we expected. However, the time series from the consensus sequence overestimates u by a factor of over thirty, and the random individual time series overestimates by a factor of a thousand. It is clear that the first two methods are inconsistent.

To see why these estimates are wrong, consider Figure 2(a), which shows the genealogical tree generated by tracing the ancestors of each gene copy back through time until the LCA. Consider, initially, two randomly chosen gene copies in the *same* generation. The probability that they have the same parent in the previous generation is $1/2N$, and the probability that they have different parents is $1 - 1/2N$. Hence, the probability that their common ancestor existed T generations in the past is

$$F(T) = \left(1 - \frac{1}{2N}\right)^{T-1} \frac{1}{2N} \approx \frac{1}{2N} \exp(-T/2N), \tag{7}$$

where the approximation assumes that $N \gg 1$. From this, the mean time back to the common ancestor is $2N$ generations. Now consider two randomly chosen gene copies at two successive generations, as in the time series method. These are the black dots in Figure 2(a). These are usually not direct ancestors of one another. The length of the chain connecting them is $2T + 1$ generations, and the 1 is negligible compared to the $2T$.

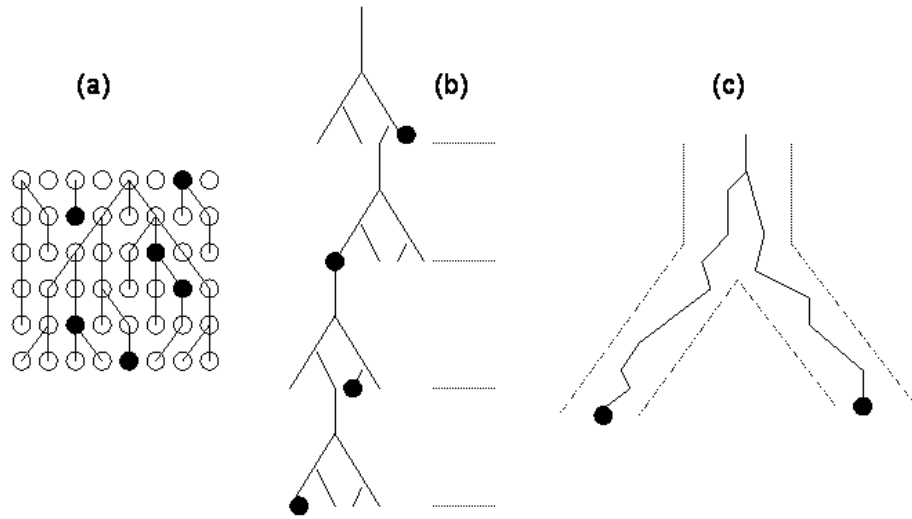


Fig. 2. (a) The lines of descent of gene copies in a population over a few generations. The black dots indicate a randomly sampled individual at each generation. (b) The relationship between gene copies sampled at large intervals from a population. (c) Illustration of a gene tree within a species tree.

For this reason, the probability that they are different is much larger than if they were direct ancestors. By combining equations 5 and 7, the probability that these two gene copies are different is

$$\int_0^\infty D(2T)F(T)dT = \frac{3}{4}\frac{16Nu}{1+16Nu}. \tag{8}$$

In our case, $16Nu = 0.1$, and this probability is 0.068, which is exactly what we obtained for the observed $\hat{D}$ in line 1 of Table 1. If this value of $\hat{D}$ is used inappropriately in equation 6, an incorrect estimate of u is obtained.

The estimate of u from the consensus sequence is also very inaccurate. This is because, when a neutral mutation spreads through a population, its frequency does not necessarily increase in a steady sweep. The frequency of the new base may wander up and down past the 50% mark several times during the transition period before it eventially goes to fixation. Therefore, the number of changes of the consensus base is more than the number of fixations. If the consensus base changed in the recent past, it is probable that the population is polymorphic, and it is quite likely that the consensus will change back to the old consensus base in the near future. On the other hand, if the consensus base did not change for a long time, the population is probably monomorphic and the consensus base is much less likely to change in the near future. Thus, the probability of change of the consensus base depends on past history, so the consensus base changes are not well described by a Markov model.

The situation is much better, however, if we only consider samples taken at broadly spaced intervals of time, rather than at every generation. Table 1 shows what happens

if we sample at intervals of $t = 10^3$ and 10^4 generations from the same simulation as before. S_{tot} is now the simulation length divided by the interval. $\hat{D}$ is larger than before because the time between each sample is larger; hence the estimates of ut are also larger. However, after dividing by t, the estimates of u are better than before. When $t = 10^3$, the estimate from the random invidiual is only out by a factor of 2, and when $t = 10^4$, it is out by less than 10%. The estimate from counting fixations is consistent, whatever the sampling interval. The estimate from the consensus sequence is between that of the other two methods.

The typical time to coalescence of all the lines of descent in a population of $2N$ gene copies can be shown to be $4N$. The interval of 10^3 in Table 1 is equal to $4N$, and the interval is 10^4 is much larger than this. When the interval is large, the relationship between the sampled sequences looks like Figure 2(b). The genealogies of the populations at each of the sampling times are now separate from one another. The LCA of each population is descended from a random individual in the previous population. Only the branches of the tree that link to individuals at the times of sampling are drawn in the figure. The population is constant in size: the figure does not imply that the population passes through repeated bottlenecks. If the sampling interval is much larger than the coalescence time, most of the changes between two successive samples occur on the line separating the populations rather than within the genealogy of each population. For this reason the different estimates of the substitution rate all converge.

Let us recap: we asked whether it is valid to represent the whole population by a single sequence and treat model it as though it were a Markov substitution process, such as the JC model. If we use either the random individual method or the consensus sequence method to generate the time series, the straightforward answer is 'no'. It is not possible to tell the probability of the nucleotide state at time $t + 1$ simply from the state at time t. Further information about the nucleotide frequencies in the population and the genealogical structure of the population is required. Therefore the assumptions of the Markov model do not hold. However, the fixation process is fairly well described by a Markov process when the allele frequency distributions are U-shaped. Hence, the method of counting fixations gives a very good estimate. When dealing with real data, we do not have a complete history of the fixation process through time; nevertheless, if we sample the population at intervals wider than the coalescence time of the population, using either a random individual or the consensus sequence, the process is quite closely described by a Markov model and the estimate of the substitution rate is very close to the true rate of fixations. Therefore, it seems that approximating the population by a single sequence might not be such a bad idea after all, which is comforting, because phylogeneticists have been doing this all along!

4. The simplest tree. Of course, the sequences used in phylogenetics do not consist of samples taken at broad intervals of time from a single lineage. Instead they consist of samples from separate species at the current time. Furthermore, we have many sites in a sequence, not just one. However, if a Markov substitution model applies for a single lineage, it will also apply for species on a tree. We will show this using the simplest possible tree: just two species.

We begin with a population of $2N = 500$, each of which has a sequence of length 2000. Evolution follows the neutral Wright-Fisher model. Each sequence is copied from the sequence of a randomly chosen parent sequence at the previous generation. Mutations occur independently at each site when the sequence is copied. After 10000 generations, a speciation event occurs, as in Figure 2c. The population is split randomly into two halves of size $2N = 250$, and these then evolve independently of one another for a period of $T = 20000$ generations. On the first generation after the split, the two populations are allowed to immediately double back to $2N = 500$, after which they remain constant in size. At the end of the simulation, one random individual is chosen from each of the two populations and the matrix of observed substitutions is calculated by comparing these two sequences.

In this example, we will use a more complex model for the mutation rates due to Tamura and Nei (TN) [7], which is defined by the matrix

$$\mathbf{r} = \begin{array}{c} \\ A \\ G \\ T \\ C \end{array} \begin{array}{c} \begin{array}{cccc} A & G & T & C \end{array} \\ \begin{pmatrix} * & \kappa_1 u\pi_G & u\pi_T & u\pi_C \\ \kappa_1 u\pi_A & * & u\pi_T & u\pi_C \\ u\pi_A & u\pi_G & * & \kappa_2 u\pi_C \\ u\pi_A & u\pi_G & \kappa_2 u\pi_T & * \end{pmatrix} \end{array} . \tag{9}$$

In the TN model, π_i is the frequency of base i. The four frequencies may be all different, as long as they sum to 1. The parameter u controls the rate of mutations between purines and pyrimidines (these are known as transversions). Transitions between purines (A and G) occur a factor κ_1 times faster than transversions, while transitions between pyrimidines (T and C) occur a factor κ_2 times faster. The $*$ denotes the fact that the matrix elements on the diagonal must be equal to minus the sum of the other elements on the row. The TN model is one of a class of phylogenetic models that are time reversible, i.e. they satisfy the rule $\pi_i r_{ij} = \pi_j r_{ji}$ for all i and j. For more examples of such models, see Higgs and Attwood [6].

The TN model is usually used as a substitution rate model in phylogenetics, but here we are using it to describe the mutation process. The probability that state i in the parent mutates to state j in the offspring is $r_{ij}\delta t$, where δt is a small time step, in this case, one generation. The probability that no mutation occurs and i remains unchanged is 1 - $\sum_{j\neq i} r_{ij}\delta t$. We used the following parameters in this example: $\pi_A = \pi_T = 0.2$, $\pi_C = \pi_G = 0.3$, $\kappa_1 = 3$, $\kappa_2 = 2$, and $u = 2 \times 10^{-5}$. The observed number of sites with i in species 1 and j in species 2 is shown below.

$$\begin{array}{c} \\ A \\ G \\ T \\ C \end{array} \begin{array}{c} \begin{array}{cccc} A & G & T & C \end{array} \\ \begin{pmatrix} 147 & 118 & 40 & 58 \\ 126 & 307 & 69 & 93 \\ 42 & 68 & 168 & 89 \\ 77 & 101 & 111 & 386 \end{pmatrix} \end{array} . \tag{10}$$

We will now use the TN model in its usual way as a substitution model for fitting the data. The object is to show that the parameters estimated from the data are consistent

with the known values used in the simulation. An analytical solution of the TN model is possible [7], by first solving equation 4 to obtain the substituion probabilities $P_{ij}(t)$, as a function of the rate parameters in equation 9, and then inverting these equations to give estimates of the rate parameters. Parameter estimates depend on the observed numbers of substitutions of three types: $\hat{D}_1$, the fraction of sites that differ by a purine transition; $\hat{D}_2$, the fraction that differ by a pyrimidine transition; and $\hat{D}_3$, the fraction that differ by a transversion. In section 3, we used $\hat{D}$ to denote the fraction of times that two successive states in the time series were different. Here, we are using $\hat{D}$ to denote the fraction of homologous sites in the two species that are different. These two quantities are analogous. It is necessary to separate the observed substitutions into three types because they depend on the rate parameters in different ways. After some algebra, the parameter estimates are:

$$\widehat{\kappa_1 ut} = -\frac{1}{2\pi_R}\ln\left(1 - \frac{\pi_R \hat{D}_1}{2\pi_A\pi_G} - \frac{\hat{D}_3}{2\pi_R}\right), \tag{11}$$

$$\widehat{\kappa_2 ut} = -\frac{1}{2\pi_Y}\ln\left(1 - \frac{\pi_Y \hat{D}_2}{2\pi_T\pi_C} - \frac{\hat{D}_3}{2\pi_Y}\right), \tag{12}$$

$$\widehat{ut} = -\frac{1}{2}\left(1 - \frac{\pi_A\pi_G}{\pi_R^2} - \frac{\pi_C\pi_T}{\pi_Y^2}\right)\ln\left(1 - \frac{\hat{D}_3}{2\pi_R\pi_Y}\right). \tag{13}$$

In the above equations, $\pi_R = \pi_A + \pi_G$ and $\pi_Y = \pi_T + \pi_C$. Table 2 shows the estimates calculated from the results in matrix 10. In a real phylogenetic problem we do not know rates and times separately, so we must be satisfied with the relative rates in column 3 of Table 2. However, in the simulation, we know that the time since the speciation is 20000 generations. The time separating the species is, therefore, approximately twice this. However, the point of coalescence of the two individuals from which the sequences were taken is slightly before the speciation, as shown in Figure 2c. We kept track of the times of common ancestors in this simulation, and it happens that the point of coalescence is 97 generations before the speciation. Therefore, the time separating the sequences is $2 \times 20097 = 40194$. Using this value of t, we obtain the absolute values of the rates in the final column of the table. These are close to the real values given above, within statistical error.

Table 2. Estimates of rate parameters in a simulation of the TN model

Type of substitution	Observed frequency	Relative rate	Absolute rate
Transitions between purines	$\hat{D}_1 = 244/2000 = 0.122$	$\kappa_1 ut = 2.58$	$\kappa_1 u = 6.4 \times 10^{-5}$
Transitions between pyrimidines	$\hat{D}_2 = 200/2000 = 0.100$	$\kappa_2 ut = 1.50$	$\kappa_2 u = 3.7 \times 10^{-5}$
Transversions	$\hat{D}_3 = 548/2000 = 0.274$	$ut = 0.795$	$u = 2.0 \times 10^{-5}$

In this case, sampling a random member of each population leads to a consistent estimation of the rate matrix. In hindsight, this should not surprise us. This is a simulation of *neutral evoluton* where each gene copy has the same fitness. The lineages of the two sequences we used for data fitting are two typical lineages in the population. Mutations

occur on these lineages independently of any mutations occurring in other members of the population. The matrix we used for mutations is the TN matrix; therefore it is not surprising that the TN model fits the data correctly. In fact, the population makes no difference in this simulation, and we could have gotten the same answer by simulating just two sequences instead of two populations. This simulation therefore demonstrates what we knew in advance from neutral evolution theory. The other point illustrated by this simulation is that the coalescence point of the two genes is slightly before the point where the species split. In cases where two speciation events occur shortly after one another, this can lead to an effect known as lineage sorting in which the branching order of the genes is not the same as the branching order of the species (Li [5], chapter 5). However, this is not a serious issue in most phylogenetic studies as long as the time between speciation events is large compared to N generations.

5. Adding natural selection. So far, things look good for the phylogeneticists: it seems to be consistent to treat the population as though it were a single sequence. However, all the models considered up to now are neutral. If selection is included, the fixation rate is no longer the same as the mutation rate. In any model with selection, the ancestors of the surviving individuals are those that were fitter than average at the time they existed. The substitutions in the lineages leading to the sampled individuals are weighted by selection, and they are not equivalent to the underlying mutations.

Suppose that the mutation process follows the JC model, so that the four bases are equivalent under mutation, but let the bases have different fitnesses, w_i. The multinomial sampling formula still applies (equation 2), but the probabilities of the four states are now dependent on the fitnesses of each state and on the mean fitness $\overline{w}$.

$$a_i = x_i(t)(1-3u)w_i/\overline{w} + \sum_{j\neq i} x_j(t)uw_j/\overline{w}. \tag{14}$$

As a specific example, consider the case where the A base is under positive selection: $w_A = 1+s$, for some small positive selection coefficient s, and $w_i = 1$ for the other three bases. There are now three types of substitutions in this model. Substitutions between any two bases other than A should be neutral, and these should occur at rate $r = u$ as in section 2. Substitutions to A from any other base are advantageous, and these should occur at rate $r = \lambda_1 u$, where λ_1 is $2N\times$ the probability of fixation of an advantageous mutation. Substitutions from A to any other base are deleterious, and these should occur at rate $r = \lambda_2 u$, where λ_2 is $2N\times$ the probability of fixation of a deleterious mutation. These fixation probabilities are known from diffusion theory (Crow and Kimura [2], Durrett [3]):

$$\lambda_1 = \frac{2N(1-e^{-2s})}{(1-e^{-4Ns})}, \tag{15}$$

$$\lambda_2 = \frac{2N(e^{2s}-1)}{(e^{4Ns}-1)}. \tag{16}$$

From consideration of the fixation process, our expected substitution rate matrix looks

like this:

$$\mathbf{r} = \begin{array}{c} \\ A \\ G \\ T \\ C \end{array} \begin{array}{c} \begin{array}{cccc} A & G & T & C \end{array} \\ \begin{pmatrix} * & \lambda_2 u & \lambda_2 u & \lambda_2 u \\ \lambda_1 u & * & u & u \\ \lambda_1 u & u & * & u \\ \lambda_1 u & u & u & * \end{pmatrix} \end{array} . \tag{17}$$

Recall that all the mutation rates are equal to u, and that the λ factors arise from the population genetics, not directly from the mutations. This is another time reversible substitution rate matrix for which the relevant quantities can be calculated using similar methods as before. The stationary frequencies of the four bases under this model are $\pi_A = \lambda_1/(\lambda_1 + 3\lambda_2)$ and $\pi_i = \lambda_2/(\lambda_1 + 3\lambda_2)$ for $i \neq$ A. Due to selection, π_A is larger than the others. If this model is fitted to sequence data, the ratio λ_1/λ_2 can be obtained from the observed value of π_A:

$$\lambda_1/\lambda_2 = 3\pi_A/(1 - \pi_A). \tag{18}$$

It is necessary to distinguish two types of substitutions. $\hat{D}_1$ is the fraction of substitutions in which one is A and one is anything other than A, and $\hat{D}_2$ is the fraction of substitutions in which the two bases are different but neither is A. Rate parameters can then be estimated:

$$\widehat{\lambda_1 ut} = -\pi_A \ln\left(1 - \frac{\hat{D}_1}{2\pi_A(1 - \pi_A)}\right), \tag{19}$$

$$\widehat{ut} = -\frac{1}{3}\ln\left(1 - \frac{3\hat{D}_2 + \hat{D}_1}{2(1 - \pi_A)}\right) - \frac{\widehat{\lambda_1 ut}}{3}. \tag{20}$$

We simulated a population evolving according to this population genetics model with parameters $2N = 500, u = 2.5 \times 10^{-5}$, and $s = 1 \times 10^{-3}$, and constructed time series for the base state as in section 2. For these parameters, $\lambda_1 = 1.58, \lambda_2 = 0.58$, $\lambda_1/\lambda_2 = 2.71$ and the expected frequency of A is $\pi_A = 0.475$. The sampling interval was 10^4 generations; hence $ut = 0.25$ and $\lambda_1 ut = 0.395$. The results are shown in Table 3. As the sampling interval is large, the random individual and consensus state methods are close to the result from counting fixations, and all three methods give a good estimate of the substitution process. Once again, as long as we are interested in intervals of time longer than the coalescence time, it is reasonable to approximate the population by a single sequence and to fit the data with a Markov model of substitutions. However, this example makes it clear that the substitution model describes the fixation process, not the mutation process.

Table 3. Estimating the substitution rate parameters from simulation data for the single site model with selection

method	π_A	$\hat{D}_1$	$\hat{D}_2$	λ_1/λ_2	$\lambda_1 ut$	ut
random individual	0.472	0.298	0.154	2.68	0.429	0.280
consensus state	0.477	0.288	0.149	2.73	0.411	0.266
counting fixations	0.482	0.281	0.143	2.80	0.398	0.251

To make this model for selection at a single site more practical for phylogenetics, it would be necessary to generalize it to deal with a whole sequence. There is no reason that the preferred base at every site should be A, so it would be necessary to allow different sites to have different preferred bases. Also, there is no reason why the same strength of selection should apply at every site, so it would be preferrable to include some kind of distribution of selection strengths across sites. These complications go beyond what is usually done in phylogenetics, where one substitution matrix is typically used to fit the average properties of a heterogenous bunch of sites. We will not pursue this here. Instead, we will consider two cases that are relevant for practical phylogenetic studies and that illustrate the importance of population genetics in the substitution process - pairs of sites in the stem regions of RNAs and triplets of sites in codon-based models for coding regions of DNA.

6. RNA pairs and codon triplets. RNA molecules such as ribosomal RNA and transfer RNA have secondary structures that are strongly conserved over evolutionary time. Each site in a stem region of an RNA structure is paired with another site on the opposite side of the stem. These pairs of sites evolve under the constraints of base-pair complementarity. In the simplest population genetics model for a pair of interacting sites, we suppose that the four combinations that form a Watson-Crick pair (AU, UA, GC, CG) are all equally fit ($w = 1$), and that all the other combinations are less fit ($w = 1 - s$) because the secondary structure is destabilized. In real RNA sequence alignments we frequently observe pairs of species that differ by pairs of mutations in the stem regions: for example, an AU pair in one species and a GC pair in another. This requires two mutations at separate sites. As mutation rates are very low, it is unlikely that both these mutations occurred simultaneously in the same gene copy. There must have been genes in the population where the sites did not match. In this example, the mismatch state would be either GU or AC. When the second mutation occurred in the mismatched sequence, the GC matching pair would be created. These pairs of mutations are called compensatory mutations, because the second mutation compensates for the deleterious effect of the first.

If the population is initially monomorphic for the AU pair, the rate of fixation of either GU or AC mismatch states is $\lambda_2 u$ (equation 16). If the mismatch state is fixed, the rate of fixation of the compensatory mutation GC (or of return to AU) should be $\lambda_1 u$ (equation 15). From this argument, the net rate of the complementary change would be slow, because it would be dominated by the slow rate of fixation of the initial deleterious mutation. However, this arguments fails to capture the nature of the complementary substitution process. If $Ns \gg 1$, and $u/s \ll 1$, it is very unlikely that a mismatch state will be fixed in the population, but there is nevertheless a continual rate of deleterious mutations to mismatch states all the time. For a large population, the frequency of a mismatch state in a population that is dominated by a matched state is u/s, determined by mutation-selection balance. Mismatch sequences will thus be present in the population at a low level, although we will typically not see them if we only sample one individual from the population. The rate of creation of GC se-

quences in an AU population is $2N \times 2u/s \times u$, where the second 2 comes from the fact that there are both GU and AC mismatches. Although the GC sequence is advantageous with respect to the mismatch, it is neutral with respect to the majority of the population, which is still AU. The probability of fixation of the neutral GC sequence is therefore $1/2N$. Thus, the net rate of the compensatory substitution is $2u^2/s$, which is much larger than $\lambda_2 u$ in the parameter range that we are considering. A more detailed calculation of the rate of compensatory subtitutions is given by Stephan [8]. The allele frequency distribution, analagous to Figure 1 can also be calculated for this situation (Higgs [9]).

Analysis of real RNA sequence alignments shows that pairs of compensatory subsitutions are frequent, and are often more frequent than single substitutions to mismatch states (Higgs [10]). The observation that double substitutions are more frequent than single substitutions is counter-intuitive initially, but the above argument shows how it can be explained in terms of population genetics.

An important aspect of real RNA structures is that GU and UG pairs occur quite frequently, and they are usually treated as weakly matching pairs rather than as mismatches. GU and UG pairs would therefore have a smaller selective disadvantage s than other mismatch pairs. As a result, changes that are double substitutions (AU $\leftrightarrow$ GC or UA $\leftrightarrow$ CG), which can go via GU or UG intermediates, occur more frequently than double transversions (such as AU $\leftrightarrow$ UA), which can only go via true mismatches. Also the severity of a mismatch mutation depends on the context of the structure in which it occurs. Free energies of helix formation depend on stacking interactions between neighbouring pairs, not just on individual pairs. A mismatch occurring in a very strongly bonded helix might be almost neutral, because the helix would still form easily when the mismatch is present, but the equivalent mismatch occurring in a weakly bonded helix might make the helix unstable, and therefore have a much more negative effect. At sites where the mismatch is nearly neutral, it would be relatively easy for the single mismatch to be fixed. We do in fact see a substantial number of sites with mismatches (especially GU and UG), as well as sites that appear to evolve almost exclusively via compensatory pairs.

Ribosomal RNA is one of the most frequently used genes in phylogenetic studies. Therefore it is of interest to develop realistic substitution rate models for the evolution of paired sites in RNA. When fitting substitution rate models to sequence data, we will be averaging over many sites with different selection regimes. A good model must therefore allow for all the types of substitutions that might occur - in particular, it must allow both single substitutions and compensatory double substitutions as possibilities. The relative rates of these processes can be estimated by fitting to real data. Among the first to use a model of this type were Tillier and Collins [11], although a variety of previous models had been proposed that only allowed single substitutions. Savill et al. [12] compared a large number of models for paired sites and used statistical tests to determine which gave better fits to sequence data. It was found that models that allowed double substitutions always fitted the data better than those that did not, and that the estimated rate of double substitutions was relatively large.

We have since gone on to implement RNA-specific models in a software package known as PHASE, that uses the Bayesian Markov Chain Monte Carlo technique for phylogenetic inference (Jow et al. [13], Hudelot et al. [14]). In phylogenetics, we are interested in determining the relative likelihood of alternative trees, given the sequence data. Standard single-site models assume that sites evolve independently and the full likelihood is a product of likelihoods of sites. RNA-pair models account for the strong correlation between the two sites in each pair. If this is ignored, relative likelihoods of alternative trees can be seriously in error. We therefore recommend using a base pair model for phylogenetics with RNA sequences, and we also recommend using a rate model that allows double substitutions, because this is an essential part of the way compensatory substitutions are fixed in populations. As we have emphasized in this article, the substitution rate model in phylogenetics needs to describe the fixation process, not the mutation process.

We now turn to models for amino acid sequences. A 20×20 substitution rate matrix can be used to describe the rate of substitution of any amino acid to any other. The earliest of these is the PAM model (Dayhoff et al. [15]), and more recent versions include those of Jones et al. [16], Adachi and Hasegawa [17], Müller and Vingron [18], and Whelan and Goldman [19]. These papers describe several different methods by which rate parameters can be estimated by fitting sequence data. As many parameters are being estimated, a very large set of sequences is required (alignments of many genes, each containing many sequences).

These models are called empirical models, because the rates are estimated from sequence data without any prior theory of which rates should be high or low. Nevertheless, the resulting matrices show some important trends that can be understood from first principles. The amino acid sequences evolve because mutations occur in the DNA sequences that code for them. Amino acid substituions are therefore influenced by the genetic code (i.e. the mapping from codons to amino acids). Amino acid changes that can be achieved by a DNA mutation at a single codon position occur more frequently than those that require two or three DNA mutations. Substitution rates are also influenced by selection at the protein level. Natural selection is, in general, conservative. Changes between amino acids that have very different physical properties tend to be disruptive of protein structure and function; therefore, they have a sizeable selective disadvantage and they tend not to be fixed in the population. Changes between amino acids with very similar properties are much more likely to be nearly neutral with respect to protein function; therefore they are more likely to be fixed in the population. A simple demonstration of this is given by Higgs and Attwood (Figure 4.6 of [6]) using the matrix of Jones et al. [16]. The set of amino acid pairs for which substitutions occur most frequently is a subset of the amino acid substitutions that can be achieved by a single mutation, i.e. all the rapid substitutions can be achieved by a single mutation, but not all the single mutation changes are rapid. Once again, this illustrates that the substitution process is describing fixation of new variants in the population, and is dependent on selection, not just mutation.

When a DNA sequence is translated to an amino acid sequence, information is lost, because more than one codon codes for each amino acid. Therefore, it is desireable to develop substitution rate models for coding sequences that work directly at the DNA

level. Although 4-state models can describe single sites in DNA, in order to describe the effects of the genetic code and selection on proteins, we need 64-state models that treat triplets of sites in a codon as a single unit. Codon-based models are often used (Muse and Gaut [20], Yang and Nielsen [21], Yang et al. [22], Pond and Muse [23]) in which each codon can mutate to any other codon that differs at a single position. Differences in base frequences and differences between transitions and transversions can be accounted for in a similar way to the TN model (equation 9). An additional important parameter in these models is ω, the ratio of non-synonymous to synonymous substitution rates. Variants of these models allow ω to differ between genes (lower ω indicates stronger stabilizing selection), or between sites (high ω indicates unusual sites where beneficial mutations have occurred).

In these models, all non-synonymous substitutions are equivalent. The model of Goldman and Yang [24] is more interesting, in that it allows non-synonymous substitution rates to depend on the difference in the physical properties of the amino acids coded by the two codons. A physical property distance matrix due to Grantham [25] is used, and it is assumed that the substitution rate decreases exponentially with this distance. This builds in the fact that conservative amino acid substitutions occur more frequently than disruptive ones, as was observed with the empirical matrices. A similar idea is used by Higgs et al. [26] using a different distance matrix. In this case, the distance between amino acids is a weighted sum of the differences in several different physical properties. The weights are estimated from the data. This gives an indication of which physical properties have the strongest influence on substitution rates.

A further feature of the model used by Higgs et al. [26] is that it allows changes between codons that differ at two or three positions, not just those that differ at a single position (as in all the codon models mentioned previously). Parameters are included in the model that control the relative rates of double and triple changes to single changes. It is found that double and triple changes occur less frequently than single changes, but they do occur. Statistical tests show that if the rates of double and triple changes are fixed to be zero, the fit to the data is much worse than if these rates are allowed to be non-zero. We interpret this as an indication that double and triple changes are being fixed in the population at the same time, even if they did not occur simultaneously as mutations. The situation seems to be similar to that occurring in RNA pairs. There is one case in codon-based models where the analogy with compensatory substitutions in RNA is quite good. The two blocks of serine codons, UCN and AGY, are not accessible to one another by a single mutation. It is necessary to pass through an intermediate codon for cysteine or threonine. At a site where serine is the favoured amino acid, the two codon blocks can exchange via compensatory substitutions. However, in most double or triple changes, the final amino acid is different from the initial one, and the intermediate amino acid is different from either of these. The compensatory substitution model then seems less applicable, because it is not clear that the intermediate state is less fit than the initial and final states. Nevertheless, double and triple substitutions seem to be important when fitting the data. Similar conclusions have been reached by Whelan and Goldman [27] and Kosiol et al. [28]. The case of double substitutions between the two blocks of

serine codons was also studied by Averof et al. [29], who argue that these occur as a result of mutational events that affect two neighbouring bases simultaneously, rather than because of compensatory substitutions. It should also be remembered that in the empirical amino acid substitution models discussed above, all amino acid substitutions are allowed. There is no restriction that substitution rates must be zero if the amino acids are not interchangeable via a single mutation. Thus, there seems to be no need to make this restriction in codon-based models either. The bottom line in phylogenetic studies is that we want to infer which is the best tree. We are most likely to get a reliable answer for the tree if we use a model of evolution that gives a good fit to the sequence data.

7. Towards more realistic models. Maximum likelihood methods in phylogenetics require the calculation of the likelihood that a known set of sequences evolved on a proposed tree. The maximum likelihood priciple is to select the tree for which this likelihood is the largest. The likelihood calculation is facilitated by the assumption that sites evolve independently; hence the likelihood for the whole sequence is the product of the likelihood of the sites. This assumption is made for convenience and practicality: it is clearly not true. The sites in RNA pairs or in codon triplets are obviously not independent, but it is relatively easy to incorporate this by using a model with a larger number of states that treats a pair or a triplet as a single unit, as in the previous section. In reality, correlations between different parts of a sequence occur in a way that cannot be parcelled up into neat pairs and triplets. For example, in RNA, stacking interactions between neighbouring pairs are important for the stability of the helix [10][30], so each pair is influenced by its neighbours, which are influenced by their neighbours, and so on. Similar correlations exist in protein structures too, and not just between neighbours along the chain. Residues in a protein can be in close proximity spatially even if they are not close along the backbone of the sequence.

In general, the fitness of a molecule depends on the whole molecule. Sites in a molecule are linked, so selection acting at one site can influence the fate of mutations occurring at other sites. Recent work has begun to take this into account. In both RNA and protein sequence evolution, the structure of the molecule tends to be more conserved than the sequence. The structure acts as a constraint on the evolution of the sequence. In order for a sequence to fold to a given structure, that structure must be thermodynamically favourable in comparison to other structures that the sequence could form. Computational methods are available that predict the energy of a sequence when folded to a specified structure. In the RNA case, thermodynamic parameters for formation of various kinds of base pairs and loop structures are known from experimental measurement (Freier et al. [30]). These are incorporated into routines that calculate the energy of a sequence when folded to a specified secondary structure. In the protein case, separation of secondary and tertiary structure is not so simple. An amino acid interacts with its close neighbours within the tertiary structure. There is no convenient, experimentally measured energy scale for these interactions, but pseudo-energies can be calculated statistically by analyzing the contacts that occur in many known examples of protein structures (Jones [31]). The more frequent the observed contacts between two amino acids (relative to the

expectation in random sequences), the lower should be the pseudo-energy for interaction of these amino acids. It is then possible to calculate the pseudo-energy of any sequence when threaded onto the known structure of a real protein.

Given either a real energy or a pseudo energy function, the evolution of sequences on a fixed structure can be studied. This has been done for proteins (Robinson et al. [32], Rodrigue et al. [33]) and RNAs (Yu and Thorne [34], Thorne et al. [35]). It is assumed that fitness depends on energy: mutations that lower the energy are treated as advantageous and those that increase the energy are treated as deleterious. Substitution rates are calculated in these models as functions of population genetics parameters N, s and u, as in equations 15 and 16. This leads to the intriguing possibility that by fitting data in the form of single sequences per species, as in phylogenetics, it is possible to estimate parameters that are relevant to population genetics, i.e. one can do population genetics without data on within-species variation [35].

In all these studies, it is assumed that a sequence can mutate into any sequence that differs from it by one mutation. For a nucleic acid sequence of length L, there are 4^L possible sequences and $3L$ other sequences to which each sequence could mutate in a single step. Multiple substitutions could, in principle, be incorporated into this framework, but have not yet been included, for reasons of computational complexity. It can be seen that this type of model is of a different level of complexity to those considered in all the previous sections, because the number of states increases exponentially with the sequence length, whereas in the previous cases it is fixed independently of L (e.g. 4 for a single site model or 64 for a codon model). In the likelihood calculations for standard models, the quantity calculated is a sum over the likelihoods of all possible histories of sequence evolution that could have existed in the interior portions of a tree. This requires solution of equation 4, which involves determining the eigenvalues of the r_{ij} matrix. For the models discussed in this section, an alternative method is used in which possible histories of substitution along the tree are sampled using a Markov Chain Monte Carlo method. This allows model parameters to be estimated in a Bayesian framework.

8. Conclusions. Here we will summarize the central points considered in this article. Population genetics and phylogenetics are both fields in which stochastic models have been developed to explain biological data. The models used in the two fields have largely been independent of one another because they have focused on different kinds of data: within-species variation in the case of population genetics and between-species variation in the case of phylogenetics. As the volume of useful sequence data for phylogenetics has increased over the past decades, the potential for meaningful estimation of large numbers of parameters from the data has also increased. This has driven the development of ever-more-complex models in phylogenetics, supported by the observation that more complex models can often be shown to fit real data better than simpler ones.

Phylogenetic models have now reached the level of complexity where it is desirable to think about population genetics parameters in the model specification. However, this raises the main issue with which we began this paper. Phylogenetic models consider a substitution process as a Markov process going on in a single sequence. This ignores

variability among members of a population, which is the whole raison d'etre of population genetics. Is this a valid thing to do? We concluded that, in many cases, the answer is 'yes'. If mutation rates are very small and population sizes are not too large, as is very often the case, most sites in a gene sequence will be monomorphic, and differences between species will usually be larger than differences between members of the same species. A randomly sampled member of the population will then be reasonably representative of the species as a whole. We have shown that Markov substituion rate models can be used to fit data consisting of randomly sampled sequences taken from the populations of separate species, or from the lineage of one species at broad intervals of time. The substitution rates estimated in this way correspond to the rates of fixation of new sequence variants in the population. In the case of neutral evolution, the fixation rate is equal to the mutation rate, so fitting the substitution model gives a direct estimate of the mutation rates. However, in any model applied to real data, this will not be the case. The substitution matrix will give an estimate of fixation rates, which depend on selection and random drift, in addition to mutation. A clear example of the qualitative difference between mutation and fixation rates is given by compensatory mutations, which are known to be frequent in RNA and which also appear to be occurring in protein coding sequences.

When specifying the functional form of the substitution models to be used in phylogenetics, it is necessary to allow for the types of complexity that we expect to see in the fixation rates. This complexity is a consequence of the fact that there really is a population genetics process generating the sequence data that we analyze in phylogenetics.

References

[1] S. Wright, *The Theory of Gene Frequencies*, Vol. II, University of Chicago Press, Chicago, 1969.

[2] J. F. Crow and M. Kimura, *An Introduction to Population Genetics Theory*, Harper and Row, New York, 1970.

[3] R. Durrett, *Probability models for DNA sequence evolution*, Springer, New York, 2002.

[4] T. H. Jukes and C. R. Cantor, *Evolution of protein molecules*, in: Mammalian Protein Metabolism, H. N. Munro (ed.), Academic Press, New York, 1969, 21–123.

[5] W. H. Li, *Molecular Evolution*, Sinauer Associates, Sunderland Massachusetts, 1997.

[6] P. G. Higgs and T. K. Attwood, *Bioinformatics and Molecular Evolution*, Blackwell Publishing, Malden, Massachusetts, 2005.

[7] K. Tamura and M. Nei, *Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees*, Mol. Biol. Evol. 10 (1993), 512–526.

[8] W. Stephan, *The rate of compensatory evolution*, Genetics 144 (1996), 419–426.

[9] P. G. Higgs, *Compensatory neutral mutations and the evolution of RNA*, Genetica 102/103 (1998), 91–101.

[10] P. G. Higgs, *RNA secondary structure: Physical and computational aspects*, Quarterly Reviews in Biophysics 33 (2000), 199–253.

[11] E. R. M. Tillier and R. A. Collins, *Neighbor joining and maximum likelihood with RNA sequences: Addressing the interdependence of sites*, Mol. Biol. Evol. 12 (1995), 7–15.

[12] N. J. Savill, D. C. Hoyle and P. G. Higgs, *RNA sequence evolution with secondary structure constraints: Comparison of substitution rate models using maximum likelihood methods*, Genetics 157 (2001), 399–411.

[13] H. Jow, C. Hudelot, M. Rattray and P. G. Higgs, *Bayesian phylogenetics using an RNA substitution model applied to early mammalian evolution*, Mol. Biol. Evol. 19 (2002), 1591–1601.

[14] C. Hudelot, V. Gowri-Shankar, H. Jow, M. Rattray and P. G. Higgs, *RNA-based phylogenetic methods: Application to mammalian mitochondrial RNA sequences*, Mol. Phyl. Evol. 28 (2003), 241–252.

[15] M. O. Dayhoff, R. M. Schwartz and B. C. Orcutt, *A model of evolutionary change in proteins*, Atlas of Protein Sequence and Structure 5 (1978), 343–352. National Biomedical Research Foundation, Washington, DC.

[16] D. T. Jones, W. R. Taylor and J. M. Thornton, *The rapid generation of mutation data matrices from protein sequences*, CABIOS 8 (1992), 272–282.

[17] J. Adachi and M. Hasegawa, *A model of amino acid substitution in proteins encoded by mitochondrial DNA*, J. Mol. Evol. 42 (1996), 459–468.

[18] T. Müller and M. Vingron, *Modelling amino acid replacement*, J. Comp. Biol. 7 (2000), 761–776.

[19] S. Whelan and N. Goldman, *A general empirical model of protein evolution derived from multiple protein families using a maximum likelihood approach*, Mol. Biol. Evol. 18 (2001), 691–699.

[20] S. V. Muse and B. S. Gaut, *A likelihood approach for comparing synonymous and non-synonymous nucleotide substitution rates, with application to the chloroplast genome*, Mol. Biol. Evol. 11 (1994), 715–724.

[21] Z. Yang and R. Nielsen, *Synonymous and non-synonymous rate variation in nuclear genes of mammals*, J. Mol. Evol. 146 (1998), 409–418.

[22] Z. Yang, R. Nielsen, N. Goldman and A. M. Krabbe Pedersen, *Codon-substitution models for heterogeneous selection pressure at amino acid sites*, Genetics 155 (2000), 431–449.

[23] S. K. Pond and S. V. Muse, *Site to site variation of synonymous substitution rates*, Mol. Biol. Evol. 22 (2005), 2375–2385.

[24] N. Goldman and Z. Yang, *A codon-based model of nucleotide substitutions for protein-coding DNA sequences*, Mol. Biol. Evol. 11 (1994), 725–736.

[25] R. Grantham, *Amino acid difference formula to help explain protein evolution*, Science 185 (1974), 862–864.

[26] P. G. Higgs, W. Hao and G. B. Golding, *Identification of conflicting selective effects on highly expressed genes*, Evolutionary Bioinformatics 2 (2006), 1–13.

[27] S. Whelan and N. Goldman, *Estimating the frequency of events that cause multiple-nucleotide changes*, Genetics 167 (2004), 2027–2043.

[28] C. Kosiol, I. Holmes and N. Goldman, *An empirical codon model for protein sequence evolution*, Mol. Biol. Evol. 24 (2007), 1464–1479.

[29] M. Averof, A. Rokas, K. H. Wolfe and P. M. Sharp, *Evidence for a high frequency of simultaneous double-nucleotide substitutions*, Science 287 (2000), 1283–1286.

[30] S. M. Freier, R. Kierzek, J. A. Jaeger, N. Sugimoto, M. H. Caruthers, T. Nielson and D. H. Turner, *Improved free energy parameters for prediction of RNA duplex stability*, Proc. Nat. Acad. Sci. USA 83 (1986), 9373–9377.

[31] D.T. Jones, *GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences*, J. Mol. Biol. 287 (1999), 797–815.

[32] D. M. Robinson, D. T. Jones, H. Kishino, N. Goldman and J. L. Thorne, *Protein evolution with dependence among codons due to tertiary structure*, Mol. Biol. Evol. 20 (2003), 1692–1704.

[33] N. Rodrigue, N. Lartillot, D. Bryant and H. Philippe, *Site interdependence attributed to tertiary structure in amino acid sequence evolution*, Gene 347 (2005), 207–217.

[34] J. Yu and J. L. Thorne, *Dependence among sites in RNA evolution*, Mol. Biol. Evol. 23 (2006), 1527–1537.

[35] J. L. Thorne, S. C. Choi, J. Yu, P. G. Higgs and H. Kishino, *Population genetics without intraspecific data*, Mol. Biol. Evol. (in press).

STOCHASTIC MODELS IN BIOLOGICAL SCIENCES
BANACH CENTER PUBLICATIONS, VOLUME 80
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WARSZAWA 2008

MOLECULAR MOTORS AND STOCHASTIC NETWORKS

REINHARD LIPOWSKY and STEFFEN LIEPELT
Max-Planck-Institute of Colloids and Interfaces
Science Park Golm, 14424 Potsdam, Germany
E-mail: reinhard.lipowsky@mpikg-golm.mpg.de, liepelt@mpikg.mpg.de
http://www.mpikg.mpg.de/theory/

Abstract. Molecular motors are nano- or colloidal machines that keep the living cell in a highly ordered, stationary state far from equilibrium. This self-organized order is sustained by the energy transduction of the motors, which couple exergonic or 'downhill' processes to endergonic or 'uphill' processes. A particularly interesting case is provided by the chemomechanical coupling of cytoskeletal motors which use the chemical energy released during ATP hydrolysis in order to generate mechanical forces and to perform mechanical work. We describe a general network theory for molecular motors which leads to local and nonlocal balance conditions that are valid far from equilibrium and generalize the well-known detailed balance conditions for equilibrium states. The nonlocal balance conditions may also be viewed as generalizations of the classical law of mass action to small systems that transduce chemical energy into mechanical work. As a pedagogical example, we discuss the simple case of a single motor head. We also review the application to two-headed motors such as kinesin, which serves as a paradigmatic example for energy transduction via chemomechanical coupling.

1. Introduction and outline. All biological systems are built up from the same kind of building blocks. In fact, these building blocks or 'modules' form a whole hierarchy of structures that cover a wide range of length and time scales. Bottom–up, i.e., as one moves from small scales to larger ones, these structural levels are: aqueous solutions, i.e., water and ions, small molecules or monomers, polymers, supramolecular assemblies, organelles, bacterial or procaryotic cells, eucaryotic cells of animals and plants, populations of cells, tissues etc.

The most basic levels of this hierarchy correspond to nanostructures and colloids that extend from molecules to cells as shown in Fig. 1. The first level contains the small

2000 *Mathematics Subject Classification*: 92C05, 82C31, 82C35, 94C15.
Key words and phrases: stochastic processes, continuous-time Markov chains, energy transduction, entropy production.
The paper is in final form and no version of it will be published elsewhere.

 © Instytut Matematyczny PAN, 2008

molecules or monomers that provide the precursors for most macromolecules in the cell. These monomers can be divided into several subsets or 'alphabets'. Different elements of one alphabet can be connected by covalent bonds which leads to the formation of four classes of polymers: nucleic acids (RNA and DNA), proteins, polysaccharides, and lipids, see second level (or column) of Fig. 1.

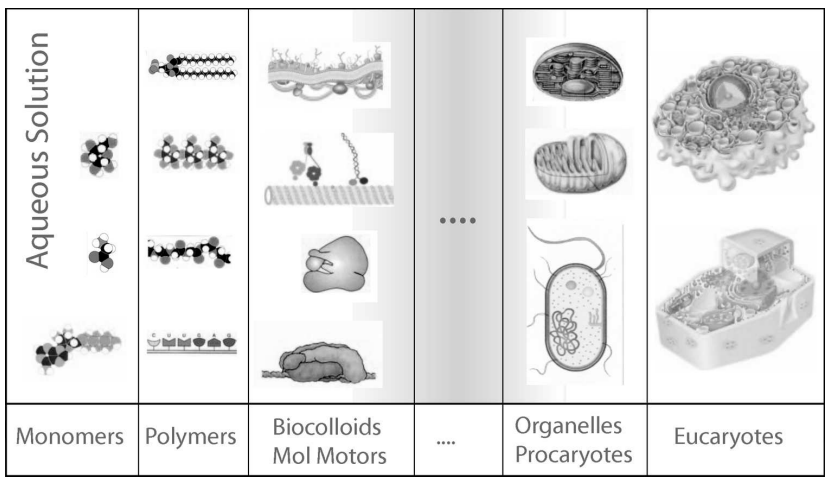


Fig. 1. Hierarchy of biological nanostructures and colloids: The first three levels on the left correspond to aqueous solutions, polymers (nucleic acids, proteins, polysaccherides, and lipids) as well as biocolloids and molecular motors (polymerases, ribosomes, cytoskeletal filaments and motors, membranes), respectively. The last two levels on the right correspond to bacterial or procaryotic cells and to eucaryotic cells. The purple regime between biocolloids and procaryotes represents the twilight zone between matter and life.

These polymers assemble into biocolloids and molecular machines, see third level of Fig. 1. The four examples shown for this latter level are in ascending order (i) a RNA polymerase that moves along a DNA strand and transcribes it into RNA, (ii) a ribosome, (iii) a cytoskeletal filament with two motor proteins, and (iv) a biomembrane. The integration of different types of biocolloids and molecular machines leads to the simplest form of life, i.e., to bacterial or procaryotic cells that have no interior subcompartments. The corresponding level in Fig. 1 also displays certain organelles such as mitochondria and chloroplasts, which are believed to have evolved from procaryotic precursor cells. Finally, the combination and integration of different types of organelles leads to eucaryotic cells as shown in the last level (or column) of Fig. 1.

Biopolymers are dead objects whereas procaryotic cells represent the simplest form of life. Thus, the regime between polymers and bacteria corresponds to the twilight zone between matter and life as indicated by the purple regime in Fig. 1. At the lower boundary of this zone, we find many different molecular motors and colloidal machines that perform various tasks such as assembly of polymers, creation of concentration gradients across membranes, transport of vesicles, cell division, and cell locomotion. These motors and machines act like little demons or nanorobots that keep the living cell in a highly ordered,

stationary state far from equilibrium, see Fig. 2(b). This self-organization is based on the energy transduction of the motors, which couple exergonic or 'downhill' processes to endergonic or 'uphill' processes. A particularly interesting case is provided by energy transduction based on chemomechanical coupling. In this latter case, the motors use the chemical energy released during ATP hydrolysis in order to generate mechanical forces and to perform mechanical work.

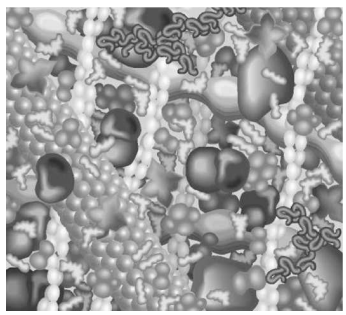
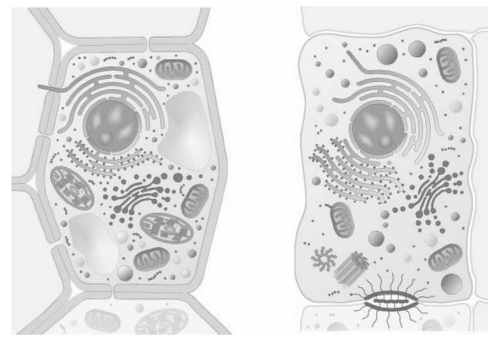
(a)  (b) 

Fig. 2. Spatial disorder and order at the molecular and colloidal scale, respectively: (a) Nanoscopic view into a living cell which reveals many different molecules and molecular assemblies that appear to be spatially disordered. The width of this image corresponds to 125 nm; and (b) Microscopic view of eucaryotic cells which exhibit soft colloidal order on the scale of hundreds of nanometers. The linear dimension of these cells is tens of micrometers.

This review is organized as follows. In Section 2, the main features of molecular motors are discussed in a qualitative manner. It is emphasized that molecular motors are colloidal machines that (i) either involve a molecular complex that interacts with a much larger supramolecular assembly or (ii) are based on the supramolecular assembly process itself. One class of motors that has been studied in much detail during the last decade consists of cytoskeletal motors such as kinesin that walk along cytoskeletal filaments by transforming the chemical energy of ATP hydrolysis into mechanical work. The thermodynamics of the motor system and its ATP hydrolysis is discussed in Section 3. As explained in this latter section, the motor is viewed as a small system that is always taken to be in thermal equilibrium but may be far from chemical equilibrium with its enviroment.

The understanding of molecular motors provides a challenge since these motors are capable of chemomechanical coupling and energy transduction even though they exhibit several aspects of randomness [1]. This behavior can be addressed within a general network theory [2, 3, 4] as explained in Sections 4, 5, and 6. Section 4 introduces network models for a single motor head or catalytic domain and for two-headed motors with two catalytic domains. The steady state fluxes of these models are discussed in Section 5, the steady state entropy production and energy balance in Section 6, in which we review our recent results as obtained in [2]. Our theory represents an extension of the theoretical framework developed by T. L. Hill and coworkers for biochemical networks [5], which do not involve mechanical forces, to the case of chemomechanical coupling as present in molecular motors. In particular, the thermodynamic forces as considered by Hill and

coworkers are shown to be equal to the entropy produced during the energy transduction. The latter relation is useful since the produced entropy can be identified with the heat released by the motor, and the latter quantity is accessible to experiment, at least in principle. A detailed comparison of our network theory with other theoretical studies is provided in [4].

As described in Section 6, our theory leads to energy balance conditions that provide generalizations of the classical law of mass action to small systems that can perform mechanical work. We also discuss a local form of the balance conditions that involves the internal energies of the motor states and the entropy produced during a single transition. The latter entropy follows directly from the decomposition of the entropy production rate in the steady state and agrees with the more general proposal of U. Seifert [6] when applied to steady states. The identification of the entropy produced by the network model with the heat released by the system relies on an appropriate level of description which includes the relevant states and transitions between these states. In particular, one should include those transitions, that describe the coupling of the motor to the different reservoirs, which are essential for energy transduction. This latter aspect is illustrated in Section 7 where we discuss energy transduction of a single motor head as described (i) by a 4-state network as introduced by T. L. Hill [5] and (ii) by a reduced 3-state network. The final Section 8 gives a brief summary of the application of our theoretical framework to the cytoskeletal motor kinesin [3], which provides a case study for chemomechanical coupling.

2. Multiscale motility of molecular motors. In each living cell, there are several types of molecular motors and machines which fulfill different functions [7]. Prominent examples are: (i) DNA and RNA polymerases which move along the strands of DNA in order to replicate it and to transcribe it into RNA; (ii) ribosomes that attach to mRNA and translate the nucleotide sequence into proteins; (iii) membrane pumps which transport ions and small molecules across membranes; the resulting concentration gradients may be used in order to drive (iv) rotary motors such as the bacterial flagellar motor and the F_0F_1–ATPase which are used for cell locomotion and ATP synthesis, respectively; (v) myosins in muscles which work in ensembles and collectively displace actin filaments; (vi) cytoskeletal motors which bind to the filaments of the cytoskeleton and then walk along these filaments in a directed fashion; and (vii) Cytoskeletal filaments that generate pushing forces during their assembly and growth.

In most cases, the molecular motor consists of a protein complex that interacts with another colloidal structure such as a macromolecule, membrane, or filament. In case (vii), the molecular machine is provided by the assembly of many proteins into a filament. In all cases, the molecular motor involves the interaction of at least two different species of molecules which bind together to form a larger colloidal complex that acts as a nanomachine.

Apart from case (iv), the molecular motors just mentioned are powered by the hydrolysis of adenosine triphosphate (ATP). Thus, all of these motors represent ATPases, i.e., catalysts or enzymes for the hydrolysis of ATP. For the concentrations which prevail in

living cells, the ATP hydrolysis is strongly exergonic or 'downhill' but it is also quite slow in the absence of any enzymatic activity. The motors act as enzymes for this chemical reaction which leads to much faster hydrolysis rates. In addition, these motors are also able to transform the energy released from the ATP hydrolysis into useful work. In fact, this energy transduction occurs even at the level of *single* hydrolysis events. In this way, they consume quantized amounts of fuel.

In this article, we will focus on cytoskeletal motors such as kinesin, see Fig. 3(a), that are essential for intracellular transport, cell division, and cell locomotion [8, 9]. Three superfamilies of cytoskeletal motors have been identified: kinesins, dyneins and myosins [9, 10]. Kinesins and dyneins bind to microtubuli as shown in Fig. 3(a) whereas myosins bind to actin filaments.

The motor properties of kinesin have been determined in a long series of sophisticated experiments [11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22]. Its processive walks consist of about 100 steps and take a few seconds [11]. In its bound state, kinesin walks in a 'hand-over-hand' fashion, i.e., by alternating steps in which one head moves forward while the other one remains bound to the filament [12, 19]. Each step corresponds to a motor displacement of 8 nm corresponding to the lattice constant of the microtubule. If there is no shortage of ATP, the motor makes about 100 steps in one second which leads to a velocity of about one micrometer per second. The absolute value of this velocity is not very impressive, but relative to its size, the motor molecule moves very fast: On the macroscopic scale, its movement would correspond to an athlete who runs 200 meters in one second! This is even more surprising if one realizes that the motor moves in a very viscous and noisy environment since it steadily undergoes thermally excited collisions with a large number of water molecules.

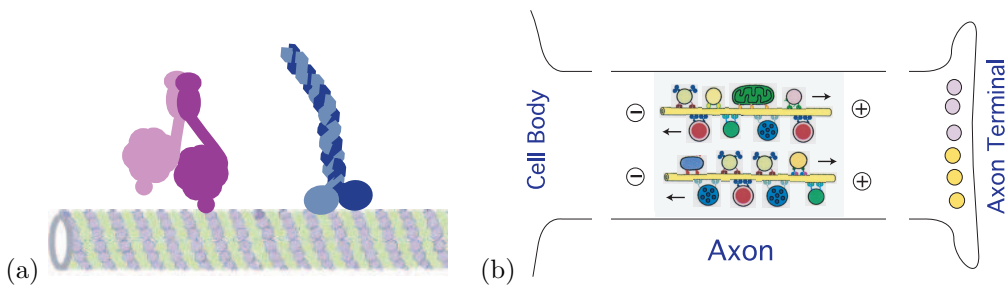


Fig. 3. Multiscale motility of molecular motors: (a) Microtubule with one dynein (violett) and one kinesin (blue) motor. The tubulin dimers form a lattice of binding sites with a lattice parameter of 8 nm; and (b) Bidirectional transport of vesicles and other types of cargo within an axon. The longest axons in our body have a length of about half a meter.

Indeed, because of this thermal noise, a single motor molecule falls off from the filament after it has made about a hundred steps. It has then covered a walking distance or run length of about one micrometer [11] which is rather small compared to the long distances – centimeters or even meters – over which cargo is transported in cells and axons, see Fig. 3(b). One rather effective way to increase the run length is via coopera-

tive transport of cargo particles by several motor molecules [23]. The corresponding run length distribution has been recently measured experimentally for an in vitro assay [24]. On length scales that are large compared to a single step along the filament, cytoskeletal motors undergo peculiar random walks that consist of both directional movement along the filament and diffusion in the surrounding solution [25, 26].

In biological cells, vesicles and other cargo particles, which are transported along microtubules, often carry both kinesins and dyneins which leads to bi-directional transport along the filaments [27, 28]. The presumably simplest mechanism for this kind of movement is provided by the competition or 'tug-of-war' between the two motor species. A recent theory for this tug-of-war leads to several coexisting motility states and is found to be consistent with the experimental data currently available [29]. In addition, many interesting transport phenomena occur if one includes the mutual exclusion between different motors or between different filaments. Indeed, both the traffic of motors on immobilized filaments [25, 26] and the traffic of filaments on immobilized motors [30, 31] leads to crowding, pattern formation, and phase transitions.

3. Thermodynamics of ATP hydrolysis. In the following, we will focus on the most basic level of motor motility as provided by its energy transduction via chemomechanical coupling. We will use the framework of statistical mechanics and view the motor as a small system that is coupled to a heat reservoir at temperature T and to particle reservoirs for the chemical species $X =$ ATP, ADP, and inorganic phosphate P. We will always assume thermal equilibrium, i.e., the motor is always characterized by the same temperature as the surrounding aqueous solution. However, the concentrations or activities of the chemical species involved in ATP hydrolysis and synthesis may or may not correspond to chemical equilibrium, which is defined as follows.

The activity of the chemical species X will be denoted by $[X]$ and has the dimension of a molar concentration. For dilute solutions, the activity is equal to the molar concentration. In general, the activity $[X]$ is defined in such a way that the chemical potential μ_X for the chemical species X has the simple form

$$\mu_X \equiv \mu_X^o + k_B T \, \ln([X]/[X]^o) \tag{3.1}$$

with Boltzmann constant k_B where the superscript o refers to some standard or reference activity $[X]^o$. The relation (3.1) implies that

$$[X]^o e^{-\mu_X^o/k_B T} = [X] e^{-\mu_X/k_B T} \equiv [X]^* \tag{3.2}$$

where the activity scale $[X]^*$ has a unique value that does not depend on the choice of the standard or reference activity $[X]^o$. Therefore, the chemical potential may also be written as

$$\mu_X = k_B T \, \ln([X]/[X]^*) \tag{3.3}$$

The activity scale $[X]^*$ can be determined using the grand-canonical ensemble. One then has to consider a large volume V of water and calculate the partition functions Z_0 and Z_1 corresponding to the situations in which this volume contains either no or a single molecule of species X, respectively. The activity scale $[X]^*$ is then given by $[X]^* = Z_1/Z_0 V N_{\rm Av}$ with the Avogardo number $N_{\rm Av}$ [32].

When the motor hydrolyzes a single ATP molecule, it binds one such molecule and releases one inorganic phosphate P and one ADP molecule. According to the Gibbs fundamental form of thermodynamics, the corresponding change in internal energy of the motor is given by

$$\Delta\mu = \mu_{\rm ATP} - \mu_{\rm P} - \mu_{\rm ADP} \tag{3.4}$$

which also represents the chemical energy input from the aqueous solution to the motor molecule. Using the expression (3.1) for the three chemical potentials, we then obtain

$$\begin{aligned} \Delta\mu &= \Delta\mu^o + k_B T \ln\!\left(\frac{[\rm ATP]}{[\rm ADP]\,[\rm P]}\frac{[\rm ADP]^o\,[\rm P]^o}{[\rm ATP]^o}\right) \\ &= k_B T \ln\!\left(\frac{[\rm ATP]}{[\rm ADP]\,[\rm P]}\frac{[\rm ADP]^*\,[\rm P]^*}{[\rm ATP]^*}\right), \end{aligned} \tag{3.5}$$

with $\Delta\mu^o \equiv \mu^o_{\rm ATP} - \mu^o_{\rm P} - \mu^o_{\rm ADP}$.

Chemical equilibrium between ATP hydrolysis and ATP synthesis corresponds to $\Delta\mu = 0$ which implies

$$\left.\frac{[\rm ADP]\,[\rm P]}{[\rm ATP]}\right|^{\rm eq} = \frac{[\rm ADP]^*\,[\rm P]^*}{[\rm ATP]^*}. \tag{3.6}$$

The right hand side of this equation defines the equilibrium (dissociation) constant $K^{\rm eq}$. Therefore, the relation (3.6) is equivalent to

$$K^{\rm eq} \equiv \left.\frac{[\rm ADP]\,[\rm P]}{[\rm ATP]}\right|^{\rm eq}. \tag{3.7}$$

For dilute solutions, the activities of the three chemical species are equal to their molar concentrations and can be directly measured (after the system has relaxed into equilibrium for a sufficiently long time). For ATP hydrolysis, the precise value of the equilibrium constant depends on the ionic conditions but a typical value is given by $K^{\rm eq} = 4.9 \times 10^{11}$ μM [33, 18].

The three activities or concentrations [ATP], [ADP], and [P] represent three control parameters that can be varied independently. Since all concentrations have to be zero or positive, they span the first octant of the three dimensional parameter space with $x = [\rm ADP]$, $y = [\rm P]$, and $z = [\rm ATP]$. The equilibrium states, which satisfy the relation (3.7), form the 2-dimensional surface as given by $z = (1/K^{\rm eq})xy$. On the other hand, all parameter values that do not lie on this surface and do not satisfy (3.7) represent states of chemical nonequilibrium.

For a molecular such as kinesin that performs mechanical work against an external load force, the complete thermodynamic description of the motor depends on the three activities or concentrations [ATP], [ADP], and [P] as well as on the force component F parallel to the filament [2, 3]. An extended discussion of this thermodynamic description is given in [4].

4. Network theory for molecular motor cycles. In the main body of this article, we will address the energy transduction of molecular motors which underlies their ability to perform mechanical work. This coupling is based on their chemomechanical motor cycles.

Such a cycle consists of a cyclic sequence of conformational transitions that enables the motor to hydrolyze ATP and to perform useful work.

Molecular motors perform their cycles in a rather noisy environment. Even in the absence of ATP, they can attain many conformational states because of thermal collisions with adjacent water molecules. Kinesin, for example, is a rather large molecule, compare Fig. 3(a), and it will bend and rotate in various ways as a result of such collisions. In addition, when bound to a filament, a motor can undergo 1-dimensional diffusion or 'passive sliding' as has been observed both for dynein at microtubules [34] and for RNA polymerase at DNA [35]. The delivery of the ATP molecules represents another stochastic process since these molecules diffuse through the surrounding solution and will have many collisions with the motor molecule before they stick to the catalytic motor domains. For kinesin, this implies that the adsorption of ATP is a random process, and that the time for this adsorption is governed by some probability distribution which depends on the ATP concentration.

In order to describe these different stochastic processes, it is necessary to construct motor models in which the motor can undergo transitions between different internal states and to incorporate the basic property that some of these transitions are coupled to ATP hydrolysis. Such a description is provided by stochastic networks as reviewed in the following. In order to be specific, we will first introduce network representations for two specific cases, namely for single motor heads and for two-headed motors. The subsequent theory is, however, rather general and can be applied to any other type of molecular motor or colloidal machine.

4.1. *Network description of single motor head.* Let us first consider a single motor head of kinesin or any other motor component with a single catalytic domain for ATP hydrolysis. In the latter process, a single molecule of ATP is broken up into a single ADP molecule and inorganic phosphate P. This process involves several substeps: adsorption of ATP, hydrolysis proper, i.e., the transformation of bound ATP into bound ADP/P, the release of P, and the release of ADP. As far as the chemical composition of the catalytic domain is concerned, one may then distinguish four different states of the motor head denoted by E, T, Θ, and D as shown in Fig. 4(a): [36] State E corresponds to the empty head, state T to the state with bound ATP, state Θ to bound ADP/P, and state D to bound ADP. Such a representation was previously used by T. L. Hill for a generic ATPase [5].

The edges between the different chemical states in Fig. 4(a) represent forward and backward transitions. The edge between state i and state j will be denoted by $\langle ij \rangle$. It consists of two directed edges or di-edges, $|ij\rangle$ and $|ji\rangle$, corresponding to the forward transition from i to j and to the backward transition from j to i, respectively. Thus, the di-edge or transition $|\mathrm{ET}\rangle$ corresponds to ATP binding to the motor head, whereas the transition $|\mathrm{TE}\rangle$ represents ATP release from this head.

Likewise, the transitions $|\Theta\mathrm{D}\rangle$, $|\mathrm{D}\Theta\rangle$, $|\mathrm{DE}\rangle$, and $|\mathrm{ED}\rangle$ describe P release, P binding, ADP release, and ADP binding, respectively. Finally, the transition $|\mathrm{T}\Theta\rangle$, corresponds to ATP hydrolysis proper and the transition $|\Theta\mathrm{T}\rangle$ to ATP synthesis from ADP and P.

The three edges $\langle \mathrm{ET} \rangle$, $\langle \Theta\mathrm{D} \rangle$, and $\langle \mathrm{DE} \rangle$ involve the binding and release of a certain molecular species from the aqueous solution. In contrast, the edge $\langle \mathrm{T}\Theta \rangle$ in the 4-state

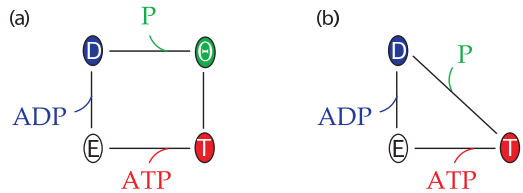


Fig. 4. Single motor head as an ATPase: (a) Network with 4 states corresponding to the motor head being empty (E), occupied by ATP (T), by ADP/P (Θ), and by ADP (D); and (b) Reduced network with 3 states in which the hydrolysis transition and the P release transition have been combined into the single transition $|31\rangle$.

network does not involve such an interaction of the motor head with the particle reservoir, see Fig. 4(a). Therefore, one may combine the two edges $\langle T\Theta\rangle$ and $\langle \Theta D\rangle$ of the 4-state network into the edge $\langle TD\rangle$ as shown in Fig. 4(b). The latter representation involves only 3 states: the motor head is occupied by ADP in state D, empty in state E, and occupied by ATP in state T. This reduced representation is useful since it eliminates some parameters but is still consistent with the fundamental constraints arising from the energetics of ATP hydrolysis, see Section 7.2 below.

4.2. *Two-headed motors with two catalytic domains.* Next, we consider a two-headed motor such as kinesin or myosin V. If each motor head can attain three different chemical states as in Fig. 4(b), the two-headed motor can attain $3^2 = 9$ different states as in Fig. 5(a). All of these states represent two-headed motors that are bound to a polar filament. We use the convention that the right head is the leading head whereas the left head is the trailing head with respect to the prefered direction of the motor movement.

If the motor walks via the 'hand-over-hand' mechanism, the leading and the trailing head interchange their positions during each *mechanical* step. If one assumes that this step is fast on the timescale of the chemical transitions, one has, in general, three possible mechanical steps: from state (E,D) to state (D,E), from state (E,T) to state (T,E), and from state (D,T) to state (T,D). These possible mechanical transitions are shown in Fig. 5(a) as broken edges. Recent experiments by Carter and Cross [21] provide strong evidence that such a separation of time scales does indeed apply to the cytoskeletal motor kinesin, see Section 8 below.

Inspection of Fig. 5(a) shows that each state of the two-headed motor is connected to four other states via four solid edges. As in Fig. 4, each solid edge between two states i and j represents both the forward chemical transition $|ij\rangle$ and the backward transition $|ji\rangle$. Thus, the 9-state network in Fig. 5(a) contains 18 solid edges corresponding to 36 *chemical* transitions.

Let us now assume that the processive motion of the two-headed motor is 'out-of-phase' and involves primarily those states for which the two heads have different chemical composition. This assumption applies to kinesin [2, 3], see section 8 below. In such a situation, one can ignore the three states (E,E), (D,D), and (T,T) in Fig. 5(a). In addition, a detailed comparison with experimental data for kinesin reveals that the only mechanical transition that is compatible with these data is provided by the transition from (D,T)

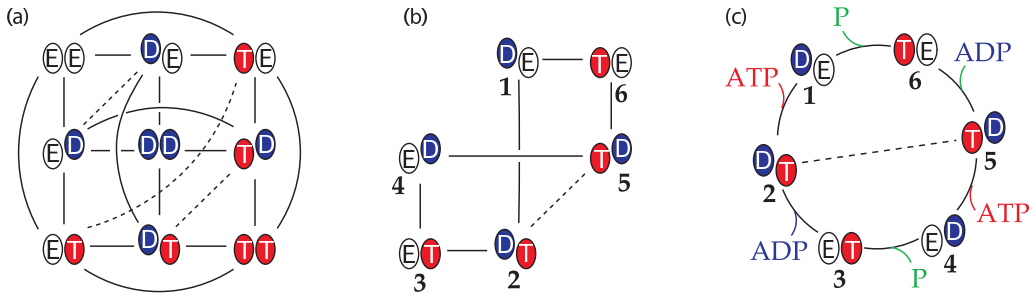


Fig. 5. Stochastic networks for a molecular motor with two catalytic domains or motor heads. Each head can be empty (E), occupied by ATP (T), or occupied by ADP (D): (a) State space with $3^2 = 9$ chemical states, each of which is connected to four neighboring states via solid edges (or lines). Each such edge represents both the forward and backward chemical transition as in Fig. 4; (b,c) Reduced state space with 6 chemical states, each of which is connected to two neighboring states via solid edges. We use the convention that the left and right head represent the trailing and leading head, respectively. Broken edges (or lines) correspond to mechanical step transitions in which the two heads interchange their relative position. In (a), three possible step transitions have been included but only one of those transitions is kept in (b) and (c). The latter transition applies to kinesin [3], see section 8 below.

to (T,D). If we delete the three states (E,E), (D,D), and (T,T) and the two mechanical transitions from (D,E) to (E,D) and from (E,T) to (T,E) from the 9-state network, we obtain the reduced 6-state network as shown in Fig. 5(b). As shown in this figure, we have labeled the six remaining states by $i = 1, 2, \ldots, 6$. A more transparent representation of the 6-state network is shown in Fig. 5(c): the six states form a cycle consisting of six (forward and backward) chemical transitions; the broken edge corresponds to the mechanical step transition. According to our convention, the mechanical forward step is given by $|25\rangle$, the mechanical backward step by $|52\rangle$.

As shown in Fig. 5(c), the cycle of chemical transitions consists (i) of the two ATP binding transitions $|12\rangle$ and $|45\rangle$, (ii) of the two transitions $|61\rangle$ and $|34\rangle$, which both represent ATP hydrolysis and P release, as well as (iii) of the two ADP release transitions $|23\rangle$ and $|56\rangle$.

4.3. *Cycles and dicycles.* The two previous subsections provided two specific examples for the description of molecular motors in terms of a discrete state space. These states are represented as the vertices of a network graph, $\mathbb{G}$, and are labeled by $i = 1, 2, \ldots, |\mathbb{G}|$. If two states i and j are connected by an edge $\langle ij \rangle$, the motor can undergo the forward transition $|ij\rangle$ from i to j and the backward transition $|ji\rangle$ from j to i. Inspection of Fig. 4 and Fig. 5 shows that these edges may form cycles. These cycles are particularly important in the present context since they are intimately related to fluxes and nonequilibrium states [5].

In order to be precise, we will distinguish (undirected) cycles from directed cycles or *dicycles*. The smallest dicycle consists of three states and three di-edges. In the mathematical literature, di-edges are often called arcs and dicycles are sometimes called circuits. An (undirected) cycle $\mathcal{C}_\nu$ is given by a closed sequence of vertices and edges in which

each vertex and each edge occurs only once. The notation $\mathcal{C}_\nu = \langle i_1 i_2 \dots i_n i_1 \rangle$ implies that the cycle $\mathcal{C}_\nu$ contains the edges $\langle i_1 i_2 \rangle$, $\langle i_2 i_3 \rangle$, ..., and $\langle i_n i_1 \rangle$. Each cycle $\mathcal{C}_\nu$ leads to two dicycles $\mathcal{C}_\nu^d$ with $d = \pm$ that differ in their orientation.

First, we choose a certain arbitrary but fixed orientation of cycle $\mathcal{C}_\nu$ to correspond to the positive direction $d = +$, which defines the dicycle $\mathcal{C}_\nu^+$. When we pass through $\mathcal{C}_\nu$ in the opposite direction $d = -$, we obtain the dicycle $\mathcal{C}_\nu^-$. These two dicycles will be denoted by $\mathcal{C}_\nu^+ = |i_1 i_2 \dots i_n i_1\rangle$ and $\mathcal{C}_\nu^- = |i_1 i_n \dots, i_2 i_1\rangle$. The dicycle $\mathcal{C}_\nu^+$ contains the di-edges $|i_1 i_2\rangle$, $|i_2 i_3\rangle$, ..., and $|i_n i_1\rangle$; the dicycle $\mathcal{C}_\nu^-$ consists of the di-edges $|i_1 i_n\rangle$, $|i_n i_{n-1}\rangle$, ..., and $|i_2 i_1\rangle$.

The network description of a single motor head, see Fig. 4, involves only a single cycle and, thus, two dicycles. Analogous unicycle models have also been frequently used for two-headed motors, see, e.g., [37, 38]. Inspection of Fig. 5 shows, however, that these motors will, in general, exhibit several motor cycles. The 9-state network in Fig. 5(a) involves a rather large number of cycles (more than 200). In contrast, the 6-state network in Fig. 5(c) contains only three cycles: the forward cycle $\mathcal{F} = \langle 25612 \rangle$, the backward cycle $\mathcal{B} = \langle 52345 \rangle$, and the chemical cycle $\mathcal{D} = \langle 1234561 \rangle$.

We will now distinguish different types of cycles by the following terminology:

(i) *Chemomechanical cycles* that contain both an ATP hydrolysis transition and a mechanical step. Examples are given by the forward cycle $\mathcal{F}$ and the backward cycle $\mathcal{B}$ in the 6-state model, see Fig. 5(c). In general, a chemomechanical cycle couples the hydrolysis of at least one ATP molecule to a net mechanical displacement;

(ii) *Dissipative slip cycles* that contain hydrolysis transitions but no mechanical step transition. One example is provided by the chemical cycle $\mathcal{D} = \langle 1234561 \rangle$ of the 6-state model;

(iii) *Mechanical slip cycles* that contain mechanical step transitions but no net hydrolysis of ATP. Examples without any hydrolysis transitions are the cycles $\langle (\mathrm{E},\mathrm{E})(\mathrm{E},\mathrm{D})(\mathrm{D},\mathrm{E})(\mathrm{E},\mathrm{E}) \rangle$ and $\langle (\mathrm{E},\mathrm{T})(\mathrm{D},\mathrm{T})(\mathrm{T},\mathrm{D})(\mathrm{T},\mathrm{E})(\mathrm{E},\mathrm{T}) \rangle$ of the 9-state model, see Fig. 5(a). An example for a mechanical slip cycle with one ATP hydrolysis and one ATP synthesis transition is provided by the cycle $\langle (\mathrm{T},\mathrm{D})(\mathrm{D},\mathrm{D})(\mathrm{D},\mathrm{T})(\mathrm{T},\mathrm{D}) \rangle$; and

(iv) *Thermal slip cycles* that involve neither a mechanical step nor a hydrolysis transition. Examples are provided by $\langle (\mathrm{E},\mathrm{E})(\mathrm{E},\mathrm{D})(\mathrm{D},\mathrm{D})(\mathrm{D},\mathrm{E})(\mathrm{E},\mathrm{E}) \rangle$ and $\langle (\mathrm{E},\mathrm{E})(\mathrm{E},\mathrm{T})(\mathrm{D},\mathrm{T})(\mathrm{D},\mathrm{E})(\mathrm{E},\mathrm{E}) \rangle$.

It is interesting to note that the reduction of the 9-state model to the 6-state model eliminates all mechanical and thermal slip cycles from the network.

5. Motor dynamics as a Markov process. The dynamics of the motor is now described by a continuous-time Markov process [39, 40] on the network graph $\mathbb{G}$. Such a process involves two stochastic ingredients, the sojourn (or dwell) times and the transition probabilities. When the system arrives in the state i, it occupies this state for a certain sojourn (or dwell) time τ_i. This time is a random variable governed by the exponential probability distribution

$$P(\tau_i) = \frac{1}{\langle \tau_i \rangle} \, e^{-\tau_i / \langle \tau_i \rangle} \tag{5.1}$$

where $\langle \tau_i \rangle$ denotes the average sojourn time [40]. When the motor leaves the state i, it jumps to state j with transition probability π_{ij}. By definition, one has $\pi_{ii} \equiv 0$ for all i and $\sum_j \pi_{ij} = 1$. From the mathematical point of view, the latter transitions occur instantaneously; from the physical point of view, they occur on a time scale that is small compared to all average sojourn times $\langle \tau_i \rangle$.

The probability $P_i(t)$ to find the motor in state i at time t is then governed by the loss-and-gain equation

$$\frac{d}{dt} P_i = -\sum_j \left(P_i \omega_{ij} - P_j \omega_{ji} \right) \tag{5.2}$$

with the transition rates

$$\omega_{ij} = \pi_{ij} / \langle \tau_i \rangle \quad \text{and} \quad \sum_j \omega_{ij} = 1 / \langle \tau_i \rangle. \tag{5.3}$$

In the physical literature, this equation is known as the Master equation [39]; in the mathematical literature, it is called the forward equation of the continuous-time Markov process [40] .

This description of the motor dynamics is convenient since it allows explicit calculation of the steady state fluxes and the steady state entropy production of the motor. These quantities are related to various experimentally accessible quantities such as the motor velocity and the ratio of forward to backward mechanical steps. In this and the following section, we will discuss the steady state fluxes and steady state entropy production, respectively.

It is convenient to define the local fluxes

$$J_{ij} \equiv P_i^{\rm st} \omega_{ij} \tag{5.4}$$

and the local excess fluxes (or currents)

$$\Delta J_{ij} \equiv P_i \omega_{ij} - P_j \omega_{ji} = -\Delta J_{ji} \tag{5.5}$$

from state i to state j. The *steady state* is characterized by probabilities $P_i^{\rm st}$ with $dP_i^{\rm st}/dt = 0$ which implies

$$\sum_j \Delta J_{ij}^{\rm st} = \sum_j (J_{ij}^{\rm eq} - J_{ji}^{\rm eq}) = \sum_j \left(P_i^{\rm st} \omega_{ij} - P_j^{\rm st} \omega_{ji} \right) = 0. \tag{5.6}$$

This relation provides $N_{\rm s}$ linear equations for the probabilities $P_i^{\rm st}$ which can be solved by linear algebra, see, e.g., [41, 42] or, more conveniently, by a graph-theoretic method [43, 44, 45, 46, 1].

The steady state equation (5.6) has a simple intuitive interpretation. The term $\sum_j J_{ij}^{\rm eq} = \sum_j P_i^{\rm st} \omega_{ij}$ represents the sum of all fluxes out of the state i; the term $\sum_j J_{ji}^{\rm eq} = \sum_j P_i^{\rm st} \omega_{ji}$ represents the sum of all fluxes into the state i. Furthermore, the local flux $J_{ij}^{\rm eq} = P_i^{\rm st} \omega_{ij}$ is equal to the number of transitions that the system performs from state i to state j per unit time in the steady state. Thus, this steady state flux is also equal to the transition frequency

$$\Omega_{ij}^{\rm st} \equiv P_i^{\rm st} \omega_{ij}, \tag{5.7}$$

i.e., to the frequency with which the system undergoes the transition $|ij\rangle$ in the steady state.

5.1. *Equilibrium, detailed balance, and dicycle ratios.* In thermal and chemical equilibrium, all local excess fluxes vanish simultaneously , i.e. $\Delta J_{ij}^{\rm eq} = 0$ for all di-edges $|ij\rangle$. In such a situation, the transition rates ω_{ij} must satisfy certain conditions, the so-called detailed balance conditions. Indeed, the relations $\Delta J_{ij}^{\rm eq} = 0$ for all di-edges can be rewritten as

$$P_i^{\rm eq}\, \omega_{ij} = P_j^{\rm eq}\, \omega_{ji} \tag{5.8}$$

for all edges $\langle ij\rangle$. These conditions can be expressed in terms of the transition rates ω_{ij} alone. In order to do so, it is convenient to define, for each dicycle $\mathcal{C}_\nu^d$ and arbitrary transition rates ω_{ij}, the *dicycle ratio*

$$\Xi(\mathcal{C}_\nu^d) \equiv \prod_{|ij\rangle \in \mathcal{C}_\nu^d} \omega_{ij} / \prod_{|ij\rangle \in \mathcal{C}_\nu^{-d}} \omega_{ij} = \prod_{|ij\rangle}^{\nu,d} (\omega_{ij}/\omega_{ji}) \tag{5.9}$$

where the superscipt ν, d at the product sign indicates a product over all di-edges $|ij\rangle$ of the dicycle $\mathcal{C}_\nu^d$. This definition implies that

$$\Xi(\mathcal{C}_\nu^-) = 1/\Xi(\mathcal{C}_\nu^+). \tag{5.10}$$

It can be shown that the detailed balance conditions as given by (5.8) are equivalent to the relations [47]

$$\Xi^{\rm eq}(\mathcal{C}_\nu^d) = \prod_{|ij\rangle}^{\nu,d} (\omega_{ij}/\omega_{ji}) = 1 \quad \text{for all dicycles } \mathcal{C}_\nu^d \tag{5.11}$$

for the dicycle ratios $\Xi^{\rm eq}(\mathcal{C}_\nu^d)$. In this form, the detailed balance conditions are no longer local (since large cycles involve widely separated edges of the network) but they are expressed only in terms of the transition rates ω_{ij}, which are the basic parameters of the model. In the context of chemical kinetics, these conditions are also known as Wegscheider conditions [48].

Away from equilibrium, some of the dicycle ratios $\Xi(\mathcal{C}_\nu^d)$ are different from one. The corresponding numerical values can be interpreted in a rather intuitive way. Indeed, as discussed in the next subsection, $\Xi(\mathcal{C}_\nu^+)$ is equal (i) to the ratio of the steady state fluxes through the dicycles $\mathcal{C}_\nu^+$ and $\mathcal{C}_\nu^-$; and (ii) to the ratio of the (unnormalized) probabilities or frequencies to complete the dicycles $\mathcal{C}_\nu^+$ and $\mathcal{C}_\nu^-$ in the steady state. In addition, we have recently shown [2] that the dicycle ratio $\Xi(\mathcal{C}_\nu^d)$ is directly related to the entropy $\Delta S(\mathcal{C}_\nu^d)$ produced during the completion of the dicycle $\mathcal{C}_\nu^d$ via the equation $\Delta S(\mathcal{C}_\nu^d) = k_B\, \ln(\Xi(\mathcal{C}_\nu^d))$, see next section.

In a long series of studies on biochemical networks, Hill and coworkers have previously discussed the quantity $k_B T\, \ln(\Xi(\mathcal{C}_\nu^d))$, which they interpreted as the "thermodynamic force" that drives the system out of equilibrium. Our results imply that this thermodynamic force is, in fact, equal to temperature T times entropy $\Delta S(\mathcal{C}_\nu^d)$ produced per completed dicycle $\mathcal{C}_\nu^d$ [2]. This identification is useful since the produced entropy can be identified with the heat released by the motor, and the latter quantity is accessible to experiment, at least in principle.

5.2. *Dicycle fluxes and frequencies.* In the steady state, each dicycle $\mathcal{C}_\nu^d$ with $d = \pm$ can be characterized by a dicycle flux $J^{\rm st}(\mathcal{C}_\nu^d)$ which represents a global property of the

stochastic network. These fluxes can be explicitly calculated (for a finite network) using a graph-theoretic or diagrammatic method. They are then found to have the general form [45, 49, 1]

$$J(\mathcal{C}_\nu^d) = \Big(\prod_{|ij\rangle}^{\nu,d} \omega_{ij}\Big) \Upsilon(\mathcal{C}_\nu)/\Omega \quad \text{with } d = \pm \tag{5.12}$$

where the superscript ν, d at the product sign indicates a product over all di-edges $|ij\rangle$ that are contained in the dicycle $\mathcal{C}_\nu^d$. The factor $\Upsilon(\mathcal{C}_\nu)$ is identical for both dicycles but depends, in general, on the cycle $\mathcal{C}_\nu$; this factor is given by a multilinear polynomial of transition rates which belong to di-edges *not* contained in the two dicycles $\mathcal{C}_\nu^+$ and $\mathcal{C}_\nu^-$. As explained in Ref. [1], these di-edges correspond to the edges of the spanning trees that are not contained in the cycle $\mathcal{C}_\nu$. The factor Ω is a normalization factor which is given by a multilinear polynomials that depends on all transition rates of the network.

Even though the dicycle fluxes $J^{\rm st}(\mathcal{C}_\nu^d)$ are global properties of the network process, they have a rather intuitive interpretation since they determine the frequencies of dicycle completion and the average completion times. Thus, let us follow the network process for a long time period Δt. The system will then pass successively through many dicycles and we can count the number of times, $N_{\Delta t}(\mathcal{C}_\nu^d)$, that it passed through the dicycle $\mathcal{C}_\nu^d$. Thus, $N_{\Delta t}(\mathcal{C}_\nu^d)$ represents the absolute frequency of completed dicycles $\mathcal{C}_\nu^d$ within the time interval Δt.

It is intuitively clear that $N_{\Delta t}(\mathcal{C}_\nu^d)$ behaves as

$$N_{\Delta t}(\mathcal{C}_\nu^d) \approx \Omega^{\rm st}(\mathcal{C}_\nu^d)\, \Delta t \quad \text{for large} \quad \Delta t \tag{5.13}$$

which defines the frequencies (or unnormalized probabilities) $\Omega^{\rm st}(\mathcal{C}_\nu^d)$. This frequency is equal to the average number of dicycles $\mathcal{C}_\nu^d$ that are completed in the steady state per unit time. It can be shown that [49, 50]

$$\Omega^{\rm st}(\mathcal{C}_\nu^d) = J^{\rm st}(\mathcal{C}_\nu^d), \tag{5.14}$$

i.e., that the dicycle completion frequencies are equal to the dicycle fluxes $J^{\rm st}(\mathcal{C}_\nu^d)$ which implies that the ratio of the dicycle completion frequencies is equal to the ratio of the dicycle fluxes. Furthermore, the explicit expression (5.12) for the dicycle fluxes implies that

$$\frac{\Omega^{\rm st}(\mathcal{C}_\nu^+)}{\Omega^{\rm st}(\mathcal{C}_\nu^-)} = \frac{J^{\rm st}(\mathcal{C}_\nu^+)}{J^{\rm st}(\mathcal{C}_\nu^-)} = \prod_{|ij\rangle}^{\nu,+} (\omega_{ij}/\omega_{ji}) = \Xi(\mathcal{C}_\nu^+) \tag{5.15}$$

i.e., both ratios are equal to the dicycle ratio $\Xi(\mathcal{C}_\nu^+)$.

The relation (5.15) for the dicycle frequencies can be generalized if one uses the action functional A of Lebowitz and Spohn [51]. Thus, consider a directed walk or trajectory $\mathcal{T}^+ \equiv |i_1 i_2 \dots i_m\rangle$ on the network graph together with the time-reversed walk $\mathcal{T}^- \equiv |i_m i_{m-1} \dots i_1\rangle$. Each walk consists of a connected sequence of di-edges $|ij\rangle$. The action functional A is then defined by [51]

$$\exp[A(\mathcal{T}^+)] \equiv \prod_{|ij\rangle \in \mathcal{T}^+} \omega_{ij} \Big/ \prod_{|ij\rangle \in \mathcal{T}^-} \omega_{ij} = \exp[-A(\mathcal{T}^-)] \tag{5.16}$$

where the product in the numerator and denominator runs over all di-edges of the walk $\mathcal{T}^+$ and $\mathcal{T}^-$, respectively.

Now, consider a closed trajectory $\mathcal{T}^+ = \mathcal{T}_\nu^+$ that goes through the dicycle $\mathcal{C}_\nu^+$ only once and does not go through any other dicycle. In general, such a trajectory consists of (i) segments of the dicycle $\mathcal{C}_\nu^+$, (ii) "back-and-forth" segments of the cycle $\mathcal{C}_\nu$ that contain segments of the dicycle $\mathcal{C}_\nu^-$ together with the corresponding reverse segments of the dicycle $\mathcal{C}_\nu^+$, and (iii) tree-like excursions away from the cycle $\mathcal{C}_\nu$. The corresponding action functional satisfies

$$\exp[A(\mathcal{T}_\nu^+)] \equiv \prod_{|ij\rangle \in \mathcal{T}_\nu^+} \omega_{ij} / \prod_{|ij\rangle \in \mathcal{T}_\nu^-} \omega_{ij} = \prod_{|ij\rangle \in \mathcal{C}_\nu^+} \omega_{ij} / \prod_{|ij\rangle \in \mathcal{C}_\nu^-} \omega_{ij} = \Xi(\mathcal{C}_\nu^+). \tag{5.17}$$

Thus, in this ratio, all transition rates from the segments of type (ii) and (iii) of the trajectories $\mathcal{T}_\nu^+$ and $\mathcal{T}_\nu^-$ cancel out, and one is simply left with the transition rates corresponding to one completion of $\mathcal{C}_\nu^+$ and $\mathcal{C}_\nu^-$. Therefore, the dicycle ratio $\Xi(\mathcal{C}_\nu^+)$ applies, in fact, to the transition rate ratios of all pairs of closed trajectories $\mathcal{T}_\nu^+$ and $\mathcal{T}_\nu^-$.

As mentioned in Ref. [51], the frequencies $\Omega^{\rm st}(\mathcal{T}^+)$ and $\Omega^{\rm st}(\mathcal{T}^-)$ for the two walks $\mathcal{T}^+$ and $\mathcal{T}^-$ are related to the action functional A via

$$\Omega^{\rm st}(\mathcal{T}^+)/\Omega^{\rm st}(\mathcal{T}^-) = (P_1^{\rm st}/P_m^{\rm st}) \exp[A(\mathcal{T}^+)]. \tag{5.18}$$

This relation simplifies for closed walks $\mathcal{T}_c^+$ which are characterized by $i_m \equiv i_1$ and, thus, by $P_1^{\rm st} = P_m^{\rm st}$. If we now consider again a closed trajectory $\mathcal{T}_c^+ = \mathcal{T}_\nu^+$ that goes through the dicycle $\mathcal{C}_\nu^+$ only once and does not go through any other dicycle, we obtain the frequency ratio

$$\frac{\Omega^{\rm st}(\mathcal{T}_\nu^+)}{\Omega^{\rm st}(\mathcal{T}_\nu^-)} = \exp[A(\mathcal{T}_\nu^+)] = \prod_{|ij\rangle}^{\nu,+} (\omega_{ij}/\omega_{ji}) = \Xi(\mathcal{C}_\nu^+). \tag{5.19}$$

As before, the superscript $\nu, +$ at the product sign indicates a product over all di-edges $|ij\rangle$ contained in the dicycle $\mathcal{C}_\nu^+$.

5.3. *Cycle decomposition of fluxes.* Using the dicyle fluxes $J(\mathcal{C}_\nu^d)$ with $d = \pm$ as in (5.12), one may define the cycle flux

$$J^{\rm st}(\mathcal{C}_\nu) \equiv J^{\rm st}(\mathcal{C}_\nu^+) - J^{\rm st}(\mathcal{C}_\nu^-) \tag{5.20}$$

which characterizes the flux through the (undirected) cycle $\mathcal{C}_\nu$. Furthermore, each excess flux ΔJ_{ij} through the di-edge $|ij\rangle$ can be decomposed according to

$$\Delta J_{ij}^{\rm st} = \sum_\nu \epsilon_{ij,\nu} J^{\rm st}(\mathcal{C}_\nu) = \sum_\nu \epsilon_{ij,\nu} \left(J^{\rm st}(\mathcal{C}_\nu^+) - J^{\rm st}(\mathcal{C}_\nu^-) \right) \tag{5.21}$$

where the antisymmetric symbol $\epsilon_{ij,\nu} = -\epsilon_{ji,\nu}$ is given by

$$\epsilon_{ij,\nu} \equiv \begin{cases} +1 & \text{for } |ij\rangle \in \mathcal{C}_\nu^+ \\ -1 & \text{for } |ij\rangle \in \mathcal{C}_\nu^- \\ 0 & \text{otherwise.} \end{cases} \tag{5.22}$$

Thus, the sum in (5.21) contains contributions from all cycles that go through the edge $\langle ij\rangle$, and the sign of these contributions is positive, if the di-edge $|ij\rangle$ is contained in (or parallel to) the dicycle $\mathcal{C}_\nu^+$, and negative, if $|ij\rangle$ is contained in (or parallel to) the reverse dicycle $\mathcal{C}_\nu^-$.

6. Entropy production and balance conditions. In this section, we will study the entropy that is produced by the motor dynamics. The entropy production rate can be decomposed in terms of separate contributions arising (i) from the different di-edges and (ii) from the different dicycles of the network graph. The dicycle entropy $\Delta S(\mathcal{C}_\nu^d)$ produced by completion of dicycle $\mathcal{C}_\nu^d$ is found to be given by $\Delta S(\mathcal{C}_\nu^d) = k_B \ln(\Xi(\mathcal{C}_\nu^d))$, i.e., by the logarithm of the dicycle ratio $\Xi(\mathcal{C}_\nu^d)$.

If one identifies the entropy produced by the motor dynamics with the heat that the motor releases into its environment, one obtains energy balance conditions that replace the detailed balance conditions for equilibrium states and hold for any state of chemical and/or mechanical nonequilibrium.

6.1. *Production of statistical or Shannon entropy.* First, we will discuss the dicycle entropies, $\Delta S(\mathcal{C}_\nu^d)$, which are produced on average, during the completion of dicycle $\mathcal{C}_\nu^d$ in the steady state, and show that these entropies satisfy the simple relations [2]

$$\Delta S(\mathcal{C}_\nu^d) \equiv k_B \ln(\Xi(\mathcal{C}_\nu^d)) = k_B \ln\Big(\prod_{|ij\rangle}^{\nu,d} (\omega_{ij}/\omega_{ji})\Big), \tag{6.1}$$

with the dicycle ratios $\Xi(\mathcal{C}_\nu^d)$. Since these ratios satisfy $\Xi(\mathcal{C}_\nu^-) = 1/\Xi(\mathcal{C}_\nu^+)$, the entropies of the two dicycles $\mathcal{C}_\nu^+$ and $\mathcal{C}_\nu^-$ are related via

$$\Delta S(\mathcal{C}_\nu^-) = -\Delta S(\mathcal{C}_\nu^+) \tag{6.2}$$

which reflects the fact that the dicycle $\mathcal{C}_\nu^-$ corresponds to the time-reversed dicycle $\mathcal{C}_\nu^+$.

T. L. Hill and coworkers have previously discussed the quantity $k_B T \ln(\Xi(\mathcal{C}_\nu^d))$, which they viewed as the "thermodynamic force" that drives the system out of equilibrium. Our work shows that this thermodynamic force is, in fact, equal to temperature T times dicycle entropy $\Delta S(\mathcal{C}_\nu^d)$.

It is interesting to note that one may combine the relations (5.15) and (6.1) in order to express the dicycle frequency ratio $\Omega^{\rm st}(\mathcal{C}_\nu^+)/\Omega^{\rm st}(\mathcal{C}_\nu^-)$ in terms of the dicycle entropies via

$$\frac{\Omega^{\rm st}(\mathcal{C}_\nu^+)}{\Omega^{\rm st}(\mathcal{C}_\nu^-)} = e^{\Delta S(\mathcal{C}_\nu^+)/k_B} = e^{-\Delta S(\mathcal{C}_\nu^-)/k_B} \tag{6.3}$$

which is reminiscent of the various relations that have been derived in the context of entropy fluctuations [52, 53, 54, 6]. However, the entropies $\Delta S(\mathcal{C}_\nu^d)$ as considered here do not fluctuate but have a certain, fixed value for each dicycle of the network. [55]

In the following, we will give two derivations of the relations (6.1) which are both based on the decomposition of the entropy production rate in the steady state. We start from the statistical or Shannon entropy

$$S\{P_i\} \equiv -k_B \sum_i P_i \ln(P_i) \tag{6.4}$$

which one may define for any probability distribution $\{P_i\}$ on a discrete state space $\mathbb{G}$. This entropy provides a well-defined measure by which one can compare different probability distributions. If the probability distribution changes with time t, so does the

statistical entropy $S\{P_i\}$. Its time derivative can be written in the form [46, 56, 51, 57]

$$\frac{d}{dt}S\{P_i\} = \sigma_{\rm pr} + \sigma_{\rm fl} \tag{6.5}$$

with the entropy production rate

$$\sigma_{\rm pr} \equiv \tfrac{1}{2}\, k_B \sum_i {\sum_j}' \Delta J_{ij}\, \ln\left(\frac{P_i\omega_{ij}}{P_j\omega_{ji}}\right) \tag{6.6}$$

and the entropy flux term

$$\sigma_{\rm fl} \equiv -\tfrac{1}{2}\, k_B \sum_i {\sum_j}' \Delta J_{ij}\, \ln\left(\frac{\omega_{ij}}{\omega_{ji}}\right) \tag{6.7}$$

where the prime at the summation sign indicates that there are no terms with $j = i$. Thus, the double sum represents a summation over all di-edges or transitions $|ij\rangle$ of the network.

In the steady state, the system's entropy does not change and $\frac{d}{dt}S\{P_i\} = 0$, which leads to the entropy production rate

$$\sigma_{\rm pr}^{\rm st} = \tfrac{1}{2} \sum_i {\sum_j}' \Delta J_{ij}^{\rm st}\; k_B \ln(\omega_{ij}/\omega_{ji}). \tag{6.8}$$

We now rewrite this expression in two different ways which leads to two different derivations of the dicycle entropy $\Delta S(\mathcal{C}_\nu^d)$ as given in (6.1). The first derivation is based on the decomposition of the entropy production rate in terms of di-edge contributions. The second derivation corresponds to our original calculation in [2] which is based on the decomposition of the entropy production rate in terms of dicycle contributions.

6.1.1. *Di-edge decomposition of entropy production rate.* In the expression (6.8) for the entropy production rate $\sigma_{\rm pr}^{\rm st}$, the di-edge or transition $|ji\rangle$ gives the same contribution as the transition $|ij\rangle$ which implies

$$\sigma_{\rm pr}^{\rm st} = \sum_{|ij\rangle} P_i^{\rm st}\omega_{ij}\; k_B \ln(\omega_{ij}/\omega_{ji}) = \sum_{|ij\rangle} \Omega_{ij}^{\rm st}\; k_B \ln(\omega_{ij}/\omega_{ji}), \tag{6.9}$$

where the sum represents a summation over all di-egdes or transitions $|ij\rangle$ of the network graph and $\Omega_{ij}^{\rm st}$ is the transition frequency as defined in (5.7). Thus, the second equation in (6.9) has the form

$$\sigma_{\rm pr}^{\rm st} = \sum_{|ij\rangle} \Omega_{ij}^{\rm st}\; \Delta S_{ij} \tag{6.10}$$

with the steady state transition entropies ΔS_{ij} defined via

$$\Delta S_{ij} \equiv k_B \ln(\omega_{ij}/\omega_{ji}) = -\Delta S_{ji}. \tag{6.11}$$

This definition is rather natural if one assumes that the entropy production rate is an additive quantity just like the entropy. Indeed, the expression (6.10) gives the total entropy production rate as a sum over contributions from all di-edges or transitions, and each of these contributions is equal to the transition frequency $\Omega_{ij}^{\rm st}$ times the transition entropy ΔS_{ij}. This conclusion agrees with the more general proposal of Seifert [6], who considered arbitrary trajectories of Markov processes. The dicycle entropy $\Delta S(\mathcal{C}_\nu^d)$ as

given by (6.1) can now be obtained from

$$\Delta S(\mathcal{C}_\nu^d) = \sum_{|ij\rangle}^{\nu,d} \Delta S_{ij}, \tag{6.12}$$

i.e., by a summation of the transition entropies ΔS_{ij} over all di-edges contained in the dicycle $\mathcal{C}_\nu^d$.

6.1.2. *Dicycle decomposition of entropy production rate.* Alternatively, let us insert the dicycle decomposition (5.21) of the local excess fluxes $\Delta J_{ij}^{\rm st}$ into the relation (6.8) for the entropy production rate $\sigma_{\rm pr}^{\rm st}$. We then obtain the expression

$$\sigma_{\rm pr}^{\rm st} = \sum_\nu J^{\rm st}(\mathcal{C}_\nu)\ \tfrac{1}{2} \sum_{i,j}{}' \epsilon_{ij,\nu}\ k_B \ln(\omega_{ij}/\omega_{ji}) \tag{6.13}$$

The summation over the di-edges may be divided up into three contributions corresponding to (i) those di-edges which are contained in the dicycle $\mathcal{C}_\nu^+$, (ii) those in $\mathcal{C}_\nu^-$, and (iii) those that are neither contained in $\mathcal{C}_\nu^+$ nor $\mathcal{C}_\nu^-$. The latter di-edges do not contribute since the antisymmetric symbol $\epsilon_{ij,\nu}$ vanishes in this case. Furthermore, the contributions from (i) and (ii) are equal which implies

$$\sigma_{\rm pr}^{\rm st} = \sum_\nu J^{\rm st}(\mathcal{C}_\nu) \sum_{|i,j\rangle}^{\nu,+} k_B \ln(\omega_{ij}/\omega_{ji}) = \sum_\nu J^{\rm st}(\mathcal{C}_\nu)\, k_B \ln(\Xi(\mathcal{C}_\nu^+)) \tag{6.14}$$

with the dicycle ratio $\Xi(\mathcal{C}_\nu^+)$ as given by (5.9). Finally, using the cycle flux decomposition $J^{\rm st}(\mathcal{C}_\nu) \equiv J^{\rm st}(\mathcal{C}_\nu^+) - J^{\rm st}(\mathcal{C}_\nu^-)$ as in (5.20) and the relation $\Xi(\mathcal{C}_\nu^-) = 1/\Xi(\mathcal{C}_\nu^+)$, one obtains

$$\sigma_{\rm pr}^{\rm st} = \sum_\nu \sum_{d=\pm} J^{\rm st}(\mathcal{C}_\nu^d)\, k_B \ln(\Xi(\mathcal{C}_\nu^+)) = \sum_\nu \sum_{d=\pm} \Omega^{\rm st}(\mathcal{C}_\nu^d)\, k_B \ln(\Xi(\mathcal{C}_\nu^+)) \tag{6.15}$$

where we have replaced the dicycle fluxes $J^{\rm st}(\mathcal{C}_\nu^d)$ by the dicycle frequencies $\Omega^{\rm st}(\mathcal{C}_\nu^d)$.

In this way, the entropy production rate $\sigma_{\rm pr}^{\rm st}$ is decomposed into separate contributions arising from all dicycles of the network. This decomposition may be rewritten in the form

$$\sigma_{\rm pr}^{\rm st} = \sum_\nu \sum_{d=\pm} \Omega^{\rm st}(\mathcal{C}_\nu^d)\, \Delta S(\mathcal{C}_\nu^d) \tag{6.16}$$

with the dicycle entropies $\Delta S(\mathcal{C}_\nu^d)$ as given by (6.1).

6.2. *Energy conservation for motor cycles.* In general, each state i of the motor represents an ensemble of molecular conformations that differ in thermally excited vibrational modes. The corresponding internal energy will be denoted by U_i. For equilibrium states, the internal energy is a state function (or state property) that does not depend on the system's history. We will now assume that we can associate such a history-independent quantity with the steady states of the motor system as well. This assumption is justified in the next section by an explicit calculation of the internal energies U_i for the network models of a single motor head as introduced in section 4.1.

In general, the internal energy U_i can change during the transition $|ij\rangle$ because of (i) energy input $E_{{\rm in},ij}$ during the transition $|ij\rangle$, (ii) work W_{ij}, which the motor performs during this transition, and (iii) heat Q_{ij}, which the motor exchanges with the reservoir.

Conservation of energy now implies

$$\Delta U_{ij} \equiv U_j - U_i = E_{\mathrm{in},ij} - W_{ij} - Q_{ij}. \tag{6.17}$$

The energy input $E_{\mathrm{in},ij}$ may arise from the coupling to an exergonic chemical reaction and/or from photon absorption. The work W_{ij} may consist of mechanical work against an external force and/or against an electrochemical potential gradient.

The energy input $E_{\mathrm{in},ij}$ may be partially *stored* in the motor molecule after the transition $|ij\rangle$, which implies $\Delta U_{ij} > 0$, and this stored energy may be transformed into work and/or heat during a later transition $|i'j'\rangle$, which implies $\Delta U_{i'j'} < 0$. However, the situation simplifies if we sum the local energy balance as given by (6.17) over any dicycle $\mathcal{C}_\nu^d$ which leads to

$$\sum_{|ij\rangle}^{\nu,d} \Delta U_{ij} = 0 = \sum_{|ij\rangle}^{\nu,d} [E_{\mathrm{in},ij} - W_{ij} - Q_{ij}] \tag{6.18}$$

or the dicycle heat

$$Q(\mathcal{C}_\nu^d) \equiv \sum_{|ij\rangle}^{\nu,d} Q_{ij} = E_{\mathrm{in}}(\mathcal{C}_\nu^d) - W(\mathcal{C}_\nu^d) \tag{6.19}$$

with the dicycle energy input $E_{\mathrm{in}}(\mathcal{C}_\nu^d)$ and the dicycle work $W(\mathcal{C}_\nu^d)$ as given by

$$E_{\mathrm{in}}(\mathcal{C}_\nu^d) \equiv \sum_{|ij\rangle}^{\nu,d} E_{\mathrm{in},ij} \quad \text{and} \quad W(\mathcal{C}_\nu^d) \equiv \sum_{|ij\rangle}^{\nu,d} W_{ij}. \tag{6.20}$$

6.3. *Nonlocal and local balance conditions.* We now want to combine the entropy produced during the motor dynamics, see subsection 6.1, with the energy conservation law as given by (6.19). In order to do so, we identify the dicycle entropy $\Delta S(\mathcal{C}_\nu^d)$ as given by (6.1) with the reduced heat $Q(\mathcal{C}_\nu^d)/T$. A combination of (6.19) and (6.1) then leads to the nonlocal balance conditions [2]

$$k_B T \ln(\Xi(\mathcal{C}_\nu^d)) = Q(\mathcal{C}_\nu^d) = E_{\mathrm{in}}(\mathcal{C}_\nu^d) - W(\mathcal{C}_\nu^d) \tag{6.21}$$

or

$$k_B T \sum_{|ij\rangle}^{\nu,d} \ln\left(\frac{\omega_{ij}}{\omega_{ji}}\right) = E_{\mathrm{in}}(\mathcal{C}_\nu^d) - W(\mathcal{C}_\nu^d) \tag{6.22}$$

between the transition rates ω_{ij}, the energy input E_{in}, and the work W for any dicycle $\mathcal{C}_\nu^d$ of the network. In equilibrium, all terms on the right hand side of these equations vanish, and we recover the detailed balance conditions $\Xi(\mathcal{C}_\nu^d) = 1$ as in (5.11).

The nonlocal balance conditions as given by (6.22) are valid for any stochastic network with an arbitrary number of cycles and extend the theoretical framework for biochemical networks as developed by T. L. Hill and coworkers [5], for which $W(\mathcal{C}_\nu^d) \equiv 0$. Note that these balance conditions relate kinetic parameters as given by the transition rates to thermodynamic controll parameters. In this sense, they represent a generalization of the law of mass action to nonequilibrium situations which involve various processes of energy transduction. Since we derived these nonlocal balance conditions from the steady state properties of the system, we called them 'steady state balance conditions' in our previous work [2, 3]. It is important to note, however, that these equations do not contain any quantities that are restricted to the steady state.

We may also derive a local form of the balance conditions if we start from the local energy balance as given by (6.17) and identify the heat Q_{ij} with the transition entropy $\Delta S_{ij} = k_B \ln(\omega_{ij}/\omega_{ji})$ as given by (6.11). This leads to the local balance conditions

$$k_B T \ln(\omega_{ij}/\omega_{ji}) = U_i - U_j + E_{\mathrm{in},ij} - W_{ij}. \tag{6.23}$$

which involve the internal energies U_i and U_j. The latter quantities should be considered as system parameters that cannot be directly controlled in an experiment. A special case of the local conditions (6.23) was obtained in Ref. [58] for simplified motor models with a single cycle and $U_i = U_j$.

The local balance conditions as given by (6.23) can be rewritten in the form

$$\frac{\omega_{ji}}{\omega_{ij}} = \frac{\exp[-U_i/k_B T]}{\exp[-U_j/k_B T]}\, e^{(-E_{\mathrm{in},ij}+W_{ij})/k_B T}. \tag{6.24}$$

As explained in [4], this relation may be interpreted as a constrained equilibrium between the motor states i and j *provided* one properly distinguishes the terms U_i and U_j, which characterize the single motor states or vertices i and j, from the terms $E_{\mathrm{in},ij}$ and W_{ij}, which are related to the transition or di-edge $|ij\rangle$.

6.4. *Practical value of balance conditions.* If one wants to apply our theory to specific motors, the nonlocal balance conditions have the advantage that they provide direct relations between two types of variables that can both be determined experimentally: the transition rates can be measured, at least in principle, via kinetic experiments whereas the energy input E_{in} and the work W represent thermodynamic control parameters. Indeed, for motors that are powered by ATP hydrolysis, the energy input can be expressed in terms of the chemical energy change $\Delta\mu$ per ATP hydrolysis as given by (3.4) and the performed work depends on the external load force F that acts parallel to the filament and on the size of the mechanical step [2, 3, 4].

Therefore, one may use the nonlocal balance conditions in two different ways. If, for a certain motor, all transition rates were known from experiments, one could use the balance conditions to estimate the experimental accuracy. In practise, some of the transition rates will be difficult to measure, and the balance conditions can then be used to estimate the values of the unknown rates. We have recently applied this latter strategy to the cytoskeletal motor kinesin [2] as summarized in section 8 below.

6.5. *Some qualifications.* The derivation of the nonlocal conditions (6.22) was based on two important physical assumptions: (i) the stochastic process that describes the motor dynamics can be represented as a continuous-time Markov process and (ii) the statistical (or Shannon) entropy produced by this process can be identified with the heat that the motor releases into its environment.

These two assumptions will only apply to a specific physical system if we chose an appropriate level of description which includes the (most) relevant states and transitions between these states. This limitation can be understood if we coarse-grain our network models. If we start from the Markov process on the 6-state model shown in Fig. 5, e.g., we can derive the effective stochastic process that governs the mechanical steps alone but this latter process is no longer Markovian [59].

Furthermore, if the statistical entropy produced in a certain network model can be identified with the heat released by the motor, this identification is lost in a coarse-grained model unless the transition rates of this coarse-grained model satisfy certain relations with the transition rates of the original model. This latter point will be illustrated in the next section for the case of a single motor head.

7. Balance conditions for single motor heads. In this section, we will consider the unicycle networks for single motor heads, which contain a single catalytic domain for ATP hydrolysis, see Fig. 4(a) and (b), and provide the presumbly simplest examples for our balance conditions. In fact, these conditions ensure that both the 4-state model and the reduced 3-state model describe the same energy transduction process. We will also show that the internal energies U_i can be expressed in terms of the transition rate constants κ_{ij} and the activity scales $[X]^*$. It turns out that analogous relations apply for network representations in general. Indeed, as shown in Ref. [4], the landscape of internal energies U_i can always be expressed in terms of the transition rate constants and the activity scales for arbitrarily complex networks.

7.1. *Balance conditions for the 4-state model.* First, let us consider again the 4-state network in Fig. 4(a). As previously explained, this network has four states $i = \mathrm{E}$, T, Θ, and D. The state E corresponds to an empty head, the state T to a head occupied by one ATP molecule, the state Θ to bound ADP/P, and the state D to bound ADP. These four states are connected by forward and backward chemical transitions with transition rates ω_{ij}. Thus, the motor head can undergo chemical transitions between the four states but does not perform any mechanical work.

Three of the chemical transitions involve the binding of a chemical species: The ATP binding transition $|\mathrm{ET}\rangle$, the P binding transition $|\mathrm{D}\Theta\rangle$, and the ADP binding transition $|\mathrm{ED}\rangle$. Therefore, the corresponding transition rates ω_{ET}, $\omega_{\mathrm{D}\Theta}$, and ω_{ED} must depend on the activities [ATP], [P], and [ADP], respectively.

The 4-state network consists of the single cycle $\mathcal{H}_{4\mathrm{s}} \equiv \langle \mathrm{ET}\Theta\mathrm{DE}\rangle$. As the motor completes the hydrolysis dicycle $\mathcal{H}_{4\mathrm{s}}^{+} = |\mathrm{ET}\Theta\mathrm{DE}\rangle$, it binds one ATP molecule during the transition $|\mathrm{ET}\rangle$, hydrolyzes this ATP into ADP/P during $|\mathrm{T}\Theta\rangle$, releases P during $|\Theta\mathrm{D}\rangle$, and releases ADP during $|\mathrm{DE}\rangle$. The corresponding chemical energy inputs $E_{\mathrm{in},ij} \equiv E_{\mathrm{ch},ij}$ are given by $E_{\mathrm{ch,ET}} = \mu_{\mathrm{ATP}}$, $E_{\mathrm{ch,T}\Theta} = 0$, $E_{\mathrm{ch},\Theta\mathrm{D}} = -\mu_{\mathrm{P}}$, and $E_{\mathrm{ch,DE}} = -\mu_{\mathrm{ADP}}$.

If we sum these energy inputs over the dicycle $\mathcal{H}_{4\mathrm{s}}^{+}$ and use the definition (3.4) of the overall chemical energy $\Delta\mu$, we obtain the dicycle energy input

$$E_{\mathrm{ch}}(\mathcal{H}_{4\mathrm{s}}^{+}) = \sum_{|ij\rangle}^{+} E_{\mathrm{ch},ij} = \Delta\mu = k_B T \ln\left(K^{\mathrm{eq}} \frac{[\mathrm{ATP}]}{[\mathrm{ADP}]\,[\mathrm{P}]}\right). \tag{7.1}$$

Since this dicycle does not perform work, the steady state balance condition for this dicycle is simply given by

$$\Xi(\mathcal{H}_{4\mathrm{s}}^{+}) = k_B T \ln\left(\frac{\omega_{\mathrm{ET}}\,\omega_{\mathrm{T}\Theta}\,\omega_{\Theta\mathrm{D}}\,\omega_{\mathrm{DE}}}{\omega_{\mathrm{TE}}\,\kappa_{\Theta\mathrm{T}}\,\kappa_{\mathrm{D}\Theta}\,\kappa_{\mathrm{ED}}}\right) = \Delta\mu. \tag{7.2}$$

This relation must hold for any values of the three activities [ATP], [ADP], and [P]. The latter requirement implies that the transition rates ω_{ET} for ATP binding, ω_{ED} for ADP

binding, and $\omega_{\mathrm{D}\Theta}$ for P binding are given by

$$\omega_{\mathrm{ET}} = \kappa_{\mathrm{ET}}\,[\mathrm{ATP}], \quad \omega_{21} = \kappa_{21}\,[\mathrm{ADP}], \quad \text{and} \quad \omega_{\mathrm{D}\Theta} = \kappa_{\mathrm{D}\Theta}\,[\mathrm{P}], \tag{7.3}$$

whereas all the other rates ω_{ij} are independent of these activities and equal to the corresponding rate constants κ_{ij}. If these dependencies are inserted into (7.2), one obtains the relation

$$\frac{\kappa_{\mathrm{ET}}\,\kappa_{\mathrm{T}\Theta}\,\kappa_{\Theta\mathrm{D}}\,\kappa_{\mathrm{DE}}}{\kappa_{\mathrm{TE}}\,\kappa_{\Theta\mathrm{T}}\,\kappa_{\mathrm{D}\Theta}\,\kappa_{\mathrm{ED}}} = K^{\mathrm{eq}} \tag{7.4}$$

between the rate constants and the equilibrium constant K^{eq}.

The relations (7.4) and (7.2) were previously derived by T. L. Hill for a generic ATPase [5]. In fact, he first derived the equilibrium relation (7.4) using detailed balance and then observed that this latter relation must remain valid even if $\Delta\mu \neq 0$. In contrast, our derivation does not use detailed balance but provides a direct derivation of the general steady state balance condition (7.2) from the energy balance of the dicycle. Detailed balance is then recovered as a special case of this general condition.

Next, we will show that the local form (6.17) of the energy balance leads to explicit expressions (i) for the internal energies U_i and (ii) for the probability distribution P_i^{eq} that the system attains in chemical equilibrium. If we apply the local energy balance relation (6.17) to the four edges of the 4-state network and identify the transition heat Q_{ij} with temperature T times the transition entropy ΔS_{ij} as given by (6.11), we obtain the internal energy changes

$$\Delta U_{\mathrm{ET}} = U_{\mathrm{T}} - U_{\mathrm{E}} = \mu_{\mathrm{ATP}} - k_B T\,\ln(\omega_{\mathrm{ET}}/\omega_{\mathrm{TE}}), \tag{7.5}$$

$$\Delta U_{\mathrm{T}\Theta} = U_{\Theta} - U_{\mathrm{T}} = -k_B T\,\ln(\omega_{\mathrm{T}\Theta}/\omega_{\Theta\mathrm{T}}), \tag{7.6}$$

$$\Delta U_{\Theta\mathrm{D}} = U_{\mathrm{D}} - U_{\Theta} = -\mu_{\mathrm{P}} - k_B T\,\ln(\omega_{\Theta\mathrm{D}}/\omega_{\mathrm{D}\Theta}), \tag{7.7}$$

and

$$\Delta U_{\mathrm{DE}} = U_{\mathrm{E}} - U_{\mathrm{D}} = -\mu_{\mathrm{ADP}} - k_B T\,\ln(\omega_{\mathrm{DE}}/\omega_{\mathrm{ED}}) \tag{7.8}$$

for the 4-state network of a single motor head.

At first sight, these energy differences ΔU_{ij} seem to depend on the chemical potentials and on the activities [ATP], [ADP], and [P]. Such a dependence is, however, not present. In order to show that all internal energy differences are independent of the various activities, we now use the definition (3.2) for the activity scales $[X]^* = [X]\exp[-\mu_X/k_B T]$ that have a certain fixed value for each chemical species X. In this way, the four relations (7.5) – (7.8) are transformed into

$$\Delta U_{\mathrm{ET}} = U_{\mathrm{T}} - U_{\mathrm{E}} = -k_B T\,\ln\left(\frac{[\mathrm{ATP}]^*\,\kappa_{\mathrm{ET}}}{\kappa_{\mathrm{TE}}}\right), \tag{7.9}$$

$$\Delta U_{\mathrm{T}\Theta} = U_{\Theta} - U_{\mathrm{T}} = -k_B T\,\ln(\kappa_{\mathrm{T}\Theta}/\kappa_{\Theta\mathrm{T}}), \tag{7.10}$$

$$\Delta U_{\Theta\mathrm{D}} = U_{\mathrm{D}} - U_{\Theta} = -k_B T\,\ln\left(\frac{\kappa_{\Theta\mathrm{D}}}{\kappa_{\mathrm{D}\Theta}\,[\mathrm{P}]^*}\right), \tag{7.11}$$

and

$$\Delta U_{\mathrm{DE}} = U_{\mathrm{E}} - U_{\mathrm{D}} = -k_B T\,\ln\left(\frac{\kappa_{\mathrm{DE}}}{\kappa_{\mathrm{ED}}\,[\mathrm{ADP}]^*}\right). \tag{7.12}$$

These expressions hold for any value of $\Delta\mu$. For the equilibrium case with $\Delta\mu = 0$, these energy differences determine the equilibrium distribution $P_i^{\rm eq}$ via the detailed balance condition $P_i^{\rm eq}\,\omega_{ij} = P_j^{\rm eq}\,\omega_{ji}$ as in (5.8). In this latter case, we can then replace the transition rate ratios ω_{ij}/ω_{ji} in the four relations (7.5) – (7.8) by the probability ratios $P_j^{\rm eq}/P_i^{\rm eq}$. As a result, we find the equilibrium distribution $\{P_i^{\rm eq}\}$ as given by

$$P_i^{\rm eq} \equiv \frac{1}{Z_{4s}}\hat{P}_i^{\rm eq} \tag{7.13}$$

with the unnormalized probabilities or Gibbs factors

$$\hat{P}_{\rm E}^{\rm eq} \equiv e^{-U_{\rm E}/k_BT}, \tag{7.14}$$

$$\hat{P}_{\rm T}^{\rm eq} \equiv e^{-(U_{\rm T}-\mu_{\rm ATP})/k_BT}, \tag{7.15}$$

$$\hat{P}_{\Theta}^{\rm eq} \equiv e^{-(U_{\Theta}-\mu_{\rm ADP}-\mu_{\rm P})/k_BT}, \tag{7.16}$$

and

$$\hat{P}_{\rm D}^{\rm eq} \equiv e^{-(U_{\rm D}-\mu_{\rm ADP})/k_BT} \tag{7.17}$$

as well as the partition function

$$Z_{4s} \equiv \hat{P}_{\rm E}^{\rm eq} + \hat{P}_{\rm T}^{\rm eq} + \hat{P}_{\Theta}^{\rm eq} + \hat{P}_{\rm D}^{\rm eq}. \tag{7.18}$$

It is important to note that we could have obtained the same distribution from a (restricted) grand canonical ensemble for the motor system which provides another consistency check of the theory.

7.2. *Balance conditions for the 3-state model.* The 4-state model in Fig. 4(a) can be reduced by combining the hydrolysis transition $|{\rm T}\Theta\rangle$ with the P release transition $|\Theta{\rm D}\rangle$ into the new, effective transition $|{\rm TD}\rangle$. As a result, one obtains the 3-state model as shown in Fig. 4(b) which now contains the single cycle $\mathcal{H}_{3s} \equiv \langle{\rm ETDE}\rangle$. The steady state balance condition for the dicycle $\mathcal{H}_{3s}^{+}$ is then given by

$$\Xi(\mathcal{H}_{3s}^{+}) = k_BT\,\ln\left(\frac{\omega_{\rm ET}\,\omega_{\rm TD}\,\omega_{\rm DE}}{\omega_{\rm TE}\,\omega_{\rm DT}\,\omega_{\rm ET}}\right) = \Delta\mu \tag{7.19}$$

We now require that the 3-state network and the 4-state network describe the same energy balance for the same value of $\Delta\mu$. Comparison of the two steady state balance conditions (7.2) and (7.19) then implies the relation

$$\frac{\omega_{\rm TD}}{\omega_{\rm DT}} = \frac{\omega_{\rm T\Theta}\,\omega_{\Theta\rm D}}{\omega_{\Theta\rm T}\,\omega_{\rm D\Theta}} \tag{7.20}$$

between the transition rates $\omega_{\rm T\Theta}$, $\omega_{\Theta\rm T}$, $\omega_{\Theta\rm D}$ and $\omega_{\rm D\Theta}$ of the 4-state network and the transition rates $\omega_{\rm TD}$ and $\omega_{\rm DT}$ of the 3-state network. This relation may also be expressed in terms of the transition entropies $\Delta S_{ij} = k_B\ln(\omega_{ij}/\omega_{ji})$ which leads to

$$\Delta S_{\rm TD} = \Delta S_{\rm T\Theta} + \Delta S_{\Theta\rm D}. \tag{7.21}$$

Furthermore, using the local balance conditions (6.23), we also find that the relation (7.20) implies

$$\Delta U_{\rm TD} = -\mu_{\rm P} - k_BT\,\ln\left(\frac{\omega_{\rm TD}}{\omega_{\rm DT}}\right) = \Delta U_{\rm T\Theta} + \Delta U_{\Theta\rm D} \tag{7.22}$$

with the energy differences $\Delta U_{\mathrm{T}\Theta}$ and $\Delta U_{\Theta\mathrm{D}}$ as given by (7.6) and (7.7) for the 4-state model.

In equilibrium with $\Delta\mu = 0$, we may again express the probabilities P_i^{eq} in terms of the internal energies U_i and the chemical potentials via the detailed balance condition $P_i^{\mathrm{eq}}\,\omega_{ij} = P_j^{\mathrm{eq}}\,\omega_{ji}$. We then recover the Gibbs factors (7.14), (7.15), and (7.17) for the unnormalized probabilities $\hat{P}_{\mathrm{E}}$, $\hat{P}_{\mathrm{T}}$, and $\hat{P}_{\mathrm{D}}$ and the equilibrium probabilities

$$P_i^{\mathrm{eq}} = \frac{1}{Z_{3s}}\,\hat{P}_i \quad \text{with} \quad Z_{3s} \equiv \hat{P}_{\mathrm{E}} + \hat{P}_{\mathrm{T}} + \hat{P}_{\mathrm{D}}. \tag{7.23}$$

Therefore, the relation (7.20) between the transition rates $\omega_{\mathrm{T}\Theta}$, $\omega_{\Theta\mathrm{T}}$, $\omega_{\Theta\mathrm{D}}$ and $\omega_{\mathrm{D}\Theta}$ of the 4-state network and the transition rates ω_{TD} and ω_{DT} of the 3-state network implies that the 3-state network is governed by the same transition entropies, internal energies, and Gibbs factors as the 4-state network. In this way, the relation (7.20) illustrates the fact that the transition rates of the coarse-grained model must satisfy certain relations with the transition rates of the original model, if we want to ensure that both models describe the same process of energy transduction. In particular, let us assume that the transition entropies $\Delta S_{\mathrm{T}\Theta}$ and $\Delta S_{\Theta\mathrm{D}}$ provide good estimates for the heat released by the motor during the transitions $|\mathrm{T}\Theta\rangle$ and $|\Theta\mathrm{D}\rangle$ in the 4-state model. We may then also identify the transition entropy ΔS_{TD} with the heat released during the transition $|\mathrm{TD}\rangle$ in the 3-state model *provided* the transition entropies satisfy the relation (7.21) which is equivalent to (7.20).

8. Kinesin: A case study for chemomechanical coupling. We have recently applied the general network theory as described in the previous sections to the cytoskeletal motor kinesin [3]. Since kinesin is a two-headed motor, we start from the network description as discussed in subsection 4.2, see Fig. 5. As mentioned, this description is based (i) on the distinct chemical states that the two-headed motor can attain and (ii) on the assumption that the mechanical steps are fast compared to the chemical transitions. Recent experimental observations by Carter and Cross [21] provide strong evidence that such a separation of time scales applies to kinesin.

In Fig. 5, we have illustrated the reduction of the 9-state network to the 6-state network, in which all motor states are characterized by two motor heads that differ in their chemical composition. A slightly different representation of this 6-state network is displayed in Fig. 6(a). As before, the network consists of 6 vertices representing the 6 possible motor states, of solid edges corresponding to the chemical forward and backward transitions, and of one broken edge for the forward and backward mechanical steps. In addition, we have now indicated the direction of the forward mechanical step by a black arrow and the direction of the ATP hydrolysis by a white double-arrow.

The 6-state network contains the two chemomechanical cycles $\mathcal{F} = \langle 12561\rangle$ and $\mathcal{B} = \langle 23452\rangle$ where the motor states are labeled as in Fig. 6(a). For small ADP concentrations, the competition of these cycles is predicted to determine the stall force of the motor at which the motor velocity vanishes. In the steady state, the motor velocity v is proportional to the local excess flux $\Delta J_{25}^{\mathrm{st}}$ through the mechanical transition $|25\rangle$ and is given by

$$v = \ell\,\Delta J_{25}^{\mathrm{st}} = \ell\,[P_2^{\mathrm{st}}\,\omega_{25} - P_5^{\mathrm{st}}\,\omega_{52}] \tag{8.1}$$

with the step size $\ell = 8$ nm. Using the cycle decomposition (5.21) for the local flux $\Delta J^{\rm st}_{25}$, one obtains

$$\begin{aligned} v/\ell &= J^{\rm st}(\mathcal{F}) - J^{\rm st}(\mathcal{B}) \\ &= J^{\rm st}(\mathcal{F}^+) - J^{\rm st}(\mathcal{F}^-) - [J^{\rm st}(\mathcal{B}^+) - J^{\rm st}(\mathcal{B}^-)]. \end{aligned} \tag{8.2}$$

Therefore, in the 6-state model, this velocity is governed by the competition between the two motor cycles $\mathcal{F}$ and $\mathcal{B}$, and vanishes for $J^{\rm st}(\mathcal{F}) = J^{\rm st}(\mathcal{B})$ which provides an implicit equation for the stall force $F_{\rm s}$. This latter equation can be rewritten in the simple form

$$\frac{\omega_{25}}{\omega_{52}} = \frac{\omega_{23}}{\omega_{56}} \quad \text{for} \quad F = F_{\rm s}. \tag{8.3}$$

This equation remains unchanged in the limit of small [ADP], in which the transition rates ω_{65} and ω_{32} and the dicycle fluxes $J^{\rm st}(\mathcal{F}^-)$ and $J^{\rm st}(\mathcal{B}^-)$ vanish. Thus, for small [ADP], the motor velocity is given by $v \approx \ell[J^{\rm st}(\mathcal{F}^+) - J^{\rm st}(\mathcal{B}^+)]$, and the stall force is determined by the two dicycles $\mathcal{F}^+$ and $\mathcal{B}^+$ which both involve ATP hydrolysis.

This prediction of our network model differs qualitatively from those of unicycle models, see, e.g., [37, 38]. Indeed, in motor models with a single motor cycle, the stall force is necessarily determined by the flux balance between ATP hydrolysis and ATP synthesis [38]. Since the ATP synthesis rate is proportional to ADP concentration, the latter balance would imply that the ATP hydrolysis rate vanishes at the stall force in the limit of small ADP concentration. The experiments by Carter and Cross [21] show, however, that the ATP hydrolysis rate is finite in this limit in agreement with our network models. As shown in [3], the 6-state network in Fig. 6(a) agrees even quantitatively with the experimental data in Ref. [21] for an appropriate choice of the transition rates.

Another quantity which can be easily calculated is the ratio, q, of forward to backward mechanical steps which is given by

$$q = \frac{P_2\,\omega_{25}}{P_5\,\omega_{52}} = 1 + \frac{v/\ell}{P_5\omega_{52}}. \tag{8.4}$$

In the 6-state network, an explicit calculation leads to

$$q = \frac{1 + \omega_{56}/\omega_{52}}{1 + \omega_{23}/\omega_{25}}. \tag{8.5}$$

If the load force F is equal to the stall force $F_{\rm s}$, one has $\omega_{25}/\omega_{52} = \omega_{23}/\omega_{56}$ as in (8.3) which implies $q = 1$. For the transition rates as determined in [3], the step ratio q as calculated for the 6-state model agrees very well with the ratio as observed experimentally in [21]. In addition, the 6-state model also agrees with the data from the Block group [16, 17] and with the data on the P-dependence of the motor velocity as observed by the Howard group in [18]. Thus, the 6-state network provides a good description for almost all measured properties of kinesin's processive motion. There is one exception, however: the decrease of motor velocity with increasing ADP concentration, as calculated for the 6-state model, is not as fast as observed experimentally in [18]. In order to match the observed velocity/ADP relationship, we found it necessary to include another motor state in which both heads contain bound ADP. The resulting 7-state network is shown in Fig. 6(b).

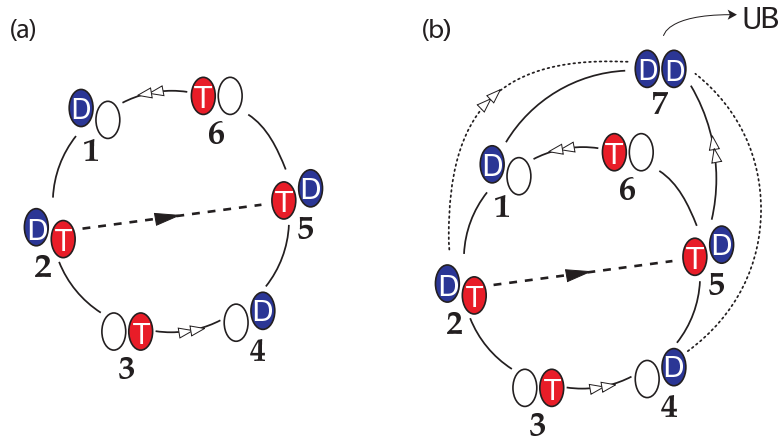


Fig. 6. Network models for kinesin: (a) 6–state network with two chemomechanical cycles, the forward cycle $\mathcal{F} = \langle 12561 \rangle$ and the backward cycle $\mathcal{B} = \langle 23452 \rangle$; and (b) 7-state network with two additional chemomechanical cycles, the additional forward cycle $\mathcal{F}_{DD} = \langle 12712 \rangle$ and the additional backward cycle $\mathcal{B}_{DD} = \langle 27452 \rangle$. In the limit of small ADP concentration, the observed dependence of the motor properties on the load force arises from the competion of the forward cycle $\mathcal{F}$ and the backward cycle $\mathcal{B}$. In the limit of small load force, on the other hand, the available experimental data on the ADP dependence of the motor velocity can be understood from the concerted action of the two forward cycles $\mathcal{F}$ and $\mathcal{F}_{DD}$ [3]. The unbinding of the motor from the filament is most likely to occur from the weakly bound state (D,D).

Inspection of Fig. 6(b) shows that the 7-state network contains two additional chemomechanical cycles, the additional forward cycle $\mathcal{F}_{DD} = \langle 12712 \rangle$ and the additional backward cycle $\mathcal{B}_{DD} = \langle 27452 \rangle$. For small load forces, the motor dynamics is found to be governed by the two forward cycles $\mathcal{F}$ and $\mathcal{F}_{DD}$. The observed reduction of the motor velocity with increasing ADP concentration can be understood as follows. First, the increase of [ADP] leads to an increase in the transition rates ω_{17} and ω_{65}, which acts to reduce the local excess fluxes $\Delta J^{\rm st}_{12} = P_1\omega_{12} - P_2\omega_{21}$ and $\Delta J^{\rm st}_{56} = P_5\omega_{56} - P_6\omega_{65}$, respectively. As a consequence, the increase of [ADP] reduces the flux $J^{\rm st}(\mathcal{F}^+)$ through the forward dicycle $\mathcal{F}^+$. This reduction of $J^{\rm st}(\mathcal{F}^+)$ is only partially compensated by the new dicycle flux $J^{\rm st}(\mathcal{F}^+_{DD})$, since this latter flux is also impeded by the more frequent transitions $|17\rangle$, which belong to the reverse dicycle $\mathcal{F}^-_{DD}$. As a result, the flux through the di-edge $|25\rangle$, which is proportional to the motor velocity, decreases substantially with increasing [ADP] as observed in experiment [3].

In summary, we have developed a network theory for the chemomechanical coupling of the molecular motor kinesin [3] that is based on the chemical state space of this motor and on the recent experimental observation [21] that its mechanical steps do not exhibit substeps. Our theory provides a quantitative description of the functional dependencies of motor velocity, ratio of forward to backward steps, and randomness parameter on the external control parameters as observed in single molecule experiments [16, 17, 18, 21].

One important aspect of our theory is the presence and competition of several chemomechanical motor cycles. For low [ADP], the motor properties are determined by the two

cycles $\mathcal{F}$ and $\mathcal{B}$, compare Fig. 6, which balance each other at the stall force. Therefore, one has to go beyond the usual unicycle models in order to understand the motor properties of kinesin. For large [ADP], yet another cycle becomes relevant that involves the weakly bound DD state. As will be described elsewhere [60], our theory can also be used to calculate additional motor properties such as the hydrolysis rate and the efficiency of the motor, which are difficult to measure experimentally.

Acknowledgements. We thank Angelo Valleriani for stimulating discussions and acknowledge support by the EC sixth Framework Program (STREP Contract No. NMP4-CT-2004-516989).

References

[1] R. Lipowsky and S. Klumpp, *Life is motion—multiscale motility of molecular motors*, Physica A 352 (2005), 53–112.

[2] S. Liepelt and R. Lipowsky, *Steady-state balance conditions for molecular motor cycles and stochastic nonequilibrium processes*, EPL 77 (2007), 50002.

[3] S. Liepelt and R. Lipowsky. *Kinesin's network of chemomechanical motor cycles*, Phys. Rev. Lett. 98 (2007), 258102.

[4] R. Lipowsky and S. Liepelt, *Chemomechanical coupling of molecular motors: Thermodynamics, Network representations, and balance conditions*, J. Stat. Phys. 130 (2008), 39–67.

[5] T. L. Hill, *Free Energy Transduction and Biochemical Cycle Kinetics*, Springer Verlag, New York, 1989.

[6] U. Seifert, *Entropy production along a stochastic trajectory and an integral fluctuation theorem*, Phys. Rev. Lett. 95 (2005), 040602/1–040602/4.

[7] B. Alberts, D. Bray, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Essential Cell Biology: An Introduction to the Molecular Biology of the Cell*, Garland, New York, 1998.

[8] J. Howard, *Mechanics of Motor Proteins and the Cytoskeleton*, Sinauer, New York, 2001.

[9] M. Schliwa and G. Woehlke, *Molecular motors*, Nature 422 (2003), 759–765.

[10] R. D. Vale, *The molecular motor toolbox for intracellular transport*, Cell 112 (2003), 467–480.

[11] S. M. Block, L. S. B. Goldstein, and B. J. Schnapp, *Bead movement by single kinesin molecules studied with optical tweezers*, Nature 348 (1990), 348–352.

[12] K. Svoboda, Ch. F. Schmidt, B. J. Schnapp, and St. M. Block, *Direct observation of kinesin stepping by optical trapping interferometry*, Nature 365 (1993), 721–727.

[13] L. Romberg and R. D. Vale, *Chemomechanical cycle of kinesin differs from that of myosin*, Nature 361 (1993), 168–170.

[14] M. J. Schnitzer and St. M. Block, *Kinesin hydrolyses one ATP per 8-nm step*, Nature 388 (1997), 386–390.

[15] S. P. Gilbert, M. L. Moyer, and K. A. Johnson, *Alternating site mechanism of the kinesin ATPase*, Biochemistry 37 (1998), 792–799.

[16] K. Visscher, M. J. Schnitzer, and S. M. Block, *Single kinesin molecules studied with a molecular force clamp*, Nature 400 (1999), 184–189.

[17] M. J. Schnitzer, K. Visscher, and S. M. Block, *Force production by single kinesin motors*, Nature Cell Biology 2 (2000), 718–723.

[18] W. R. Schief, R. H. Clark, A. H. Crevenna, and J. Howard, *Inhibition of kinesin motility by ADP and phosphate supports a hand-over-hand mechanism*, PNAS 101 (2004), 1183–1188.
[19] A. Yildiz, M. Tomishige, R. D. Vale, and P. R. Selvin, *Kinesin walks hand-over-hand*, Science 303 (2004), 676–678.
[20] D. D. Hackney, *The tethered motor domain of a kinesin–microtubule complex catalyzes reversible synthesis of bound ATP*, PNAS 102 (2005), 18338–18343.
[21] N. J. Carter and R. A. Cross, *Mechanics of the kinesin step*, Nature 435 (2005), 308–312.
[22] N. R. Guydosh and S. M. Block, *Backsteps induced by nucleotide analogs suggest the front head of kinesin is gated by strain*, Proc. Nat. Acad. Sci. USA 103 (2006), 8054–8059.
[23] S. Klumpp and R. Lipowsky, *Cooperative cargo transport by several molecular motors*, PNAS 102 (2005), 17284–17289.
[24] J. Beeg, S. Klumpp, R. Dimova, R. S. Gracia, E. Unger, and R. Lipowsky, *Transport of beads by several kinesin motors*, Biophys. J., submitted.
[25] R. Lipowsky, S. Klumpp, and T. M. Nieuwenhuizen, *Random walks of cytoskeletal motors in open and closed compartments*, Phys. Rev. Lett. 87 (2001), 108101/1–108101/4.
[26] R. Lipowsky, Y. Chai, S. Klumpp, S. Liepelt, and M. J. I. Müller, *Molecular motor traffic: From biological nanomachines to macroscopic transport*, Physica A 372 (2006), 34–51.
[27] S. P. Gross, *Hither and yon: a review of bi-directional microtubule-based transport*, Phys. Biol. 1 (2004), R1–R11.
[28] M. A. Welte, *Bidirectional transport along microtubules*, Current Biology 14 (2004), R525–R537.
[29] M. J. I. Müller, S. Klumpp, and R. Lipowsky, *Tug-of-war of molecular motors*, in preparation.
[30] P. Kraikivski, R. Lipowsky, and J. Kierfeld, *Enhanced ordering of interacting filaments by molecular motors*, Phys. Rev. Lett. 96 (2006), 258103.
[31] J. Kierfeld, P. Gutjahr, T. Kühne, P. Kraikivski, and R. Lipowsky, *Buckling, bundling, and pattern formation: From semi-flexible polymers to assemblies of interacting filaments*, J. Comp. Theo. Nanosci. 3 (2006), 898.
[32] If one dissolved the chemical species X in an ideal gas instead of water, the activity scale $[X]^*$ would be equal to $[X]^* = 1/V_X N_{\rm Av}$ with the quantum volume V_X of the species X.
[33] R. A. Alberty, *Thermodynamics of the hydrolysis of ATP as a function of temperature, pH, pMg, and ionic strength*, J. Phys. Chem. B 107 (2003), 12324–12330.
[34] R. D. Vale, D. R. Soll, and I. R. Gibbons, *One-dimensional diffusion of microtubules bound to flagellar dynein*, Cell 59 (1989), 915–925.
[35] M. Guthold, X. Zhu, C. Rivetti, G. Yang, N. G. Thomson, S. Kasas, H. G. Hansma, B. Smith, P. Hansma, and C. Bustamante, *Direct observation on one-dimensional diffusion and transcription by Escherichia coli RNA polymerase*, Biophys. J. 77 (1999), 2284–2294.
[36] Note that the symbol T is used for the motor head with bound ATP whereas the symbol T is used for the temperature.
[37] M. E. Fisher and A. Kolomeisky, *Simple mechanochemistry describes the dynamics of kinesin molecules*, Proc. Natl. Acad. Sci. USA 98 (2001), 7748–7753.
[38] M. E. Fisher and Y. C. Kim, *Kinesin crouches to sprint but resists pushing*, Proc. Natl. Acad. Sci. USA 102 (2005), 16209–16214.
[39] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, Elsevier, Amsterdam, 1992.
[40] J. R. Norris, *Markov Chains*, Cambridge University Press, Cambrigde, 1997.
[41] R. Lipowsky, *Universal aspects of the chemo-mechanical coupling for molecular motors*, Phys. Rev. Lett. 85 (2000), 4401–4404.

[42] R. Lipowsky and N. Jaster, *Molecular motor cycles: from ratchets to networks*, J. Stat. Phys. 110 (2003), 1141–1167.

[43] G. Kirchhoff, *Über die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Verteilung galvanischer Ströme geführt wird*, Ann. Phys. Chem. 72 (1847), 497–508.

[44] W.T. Tutte, *Graph Theory*, Cambridge University Press, Cambridge, 2001.

[45] T. L. Hill, *Studies in irreversible thermodynamics, IV, Diagrammatic representation of steady state fluxes for unimolecular systems*, J. Theoret. Biol. 10 (1966), 442–459.

[46] J. Schnakenberg, *Network theory of microscopic and macroscopic behavior of master equation systems*, Rev. Mod. Phys. 48 (1976), 571–585.

[47] The relation (5.11) follows from (5.8) by subsequently applying the latter relation to all di-edges of a dicycle. In addition, the relation (5.11) also implies (5.8) as follows from the flux decomposition in (5.21) and the dicycle fluxes as given by (5.12).

[48] T. A. Bak, *Contributions to the Theory of Chemical Kinetics*, Benjamin, New York, 1963.

[49] T. L. Hill and Y.-D. Chen, *Stochastics of cycle completions (fluxes) in biochemical kinetic diagrams*, Proc. Nat. Acad. Sci. USA 72 (1975), 1291–1295.

[50] H.-H. Kohler and E. Vollmerhaus, *The frequency of cyclic processes in biological multistate systems*, J. Math. Biology 9 (1980), 275–290.

[51] J. L. Lebowitz and H. Spohn, *A Gallavotti–Cohen-type symmetry in the large deviation functional for stochastic dynamics*, J. Stat. Phys. 95 (1999), 333–365.

[52] D. J. Evans, E. G. D. Cohen, and G. P. Morris, *Probability of second law violations in shearing steady states*, Phys. Rev. Lett. 71 (1993), 2401–2404.

[53] G. E. Crooks, *Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences*, Phys. Rev. E 60 (1999), 2721–2726.

[54] C. Jarzynski, *Hamiltonian derivation of a detailed fluctuation theorem*, J. Stat. Phys. 98 (2000), 77–102.

[55] In this study, the transition rates ω_{ij} are taken to be time-independent model parameters. More complex situations would arise if we allowed the transition rates to vary with time or considered ensembles of stochastic networks for which the transition rates represent fluctuating variables.

[56] J. L. Luo, C. van den Broeck, and G. Nicolis, *Stability criteria and fluctuations around non-equilibrium states*, Z. Physik B 56 (1984), 165–170.

[57] C. Maes and M. H. van Wieren, *A Markov model for kinesin*, J. Stat. Phys. 112 (2003), 329–355.

[58] U. Seifert, *Fluctuation theorem for a single enzym or molecular motor*, Europhys. Lett. 70 (2005), 36–41.

[59] A. Valleriani, S. Liepelt, and R. Lipowsky, *Waiting time distributions for kinesin's mechanical steps*, EPL (in press).

[60] S. Liepelt and R. Lipowsky, *Hydrolysis rate and efficiency of the molecular motor kinesin*, in preparation.

STOCHASTIC MODELS IN BIOLOGICAL SCIENCES
BANACH CENTER PUBLICATIONS, VOLUME 80
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WARSZAWA 2008

SPREADING MECHANISMS OF COOPERATION FOR THE EVOLUTIONARY PRISONER'S DILEMMA GAMES

GYÖRGY SZABÓ

Research Institute for Technical Physics and Materials Science
POB. 49, H-1525 Budapest, Hungary
E-mail: szabo@mfa.kfki.hu

Abstract. We survey several mechanisms supporting the maintenance of cooperation for evolutionary Prisoner's Dilemma games. In these models players are located on the sites of a lattice or graph and they can follow one of the pure strategies: cooperation (C) or defection (D). Their total income comes from Prisoner's Dilemma games with their neighbors. We discuss the consequences of different evolutionary rules determining the time-dependence of the strategy distribution and compare the results of spreading mechanisms of cooperation.

1. Introduction. Evolutionary Prisoner's Dilemma games are widely used to describe the emergence of cooperative behavior among selfish individuals (for reviews see [1-6] and further references therein). In the original Prisoner's Dilemma game two players should simultaneously choose between two options: defection (D) and cooperation (C) [7,8]. The selfish players wish to maximize their own income in the knowledge of payoffs dependent on their choices. The curiosity of Prisoner's Dilemma game comes from the fact that the choice of defection yields a higher income independent of the partner's choice. However, if both rational players choose defection then their individual income is lower than those obtained for mutual cooperation when the maximum total payoff is shared equally. The rational players cannot resolve this dilemma although we find many examples in the nature where mutual cooperation (e.g., altruism and ethical norms) emerges among the selfish individuals [9].

In the last decades several mechanisms (e.g., kin selection [10], application of retaliating strategies [11], and voluntary participation [12]) are reported which enforce the appearance of cooperation in the societies (for a recent survey see the paper by Nowak [13]). Now we will discuss two-strategy (C and D) evolutionary prisoner's dilemma games with N players ($N \to \infty$) distributed on the sites of a lattice or graph. The players play

2000 *Mathematics Subject Classification*: Primary 91A22; Secondary 91B72.
Key words and phrases: Prisoner's Dilemma games.
The paper is in final form and no version of it will be published elsewhere.

 © Instytut Matematyczny PAN, 2008

iterated games with their neighbors and sometime they are allowed to modify their strategy. More precisely, the players can adopt (learn) one of the neighboring strategies in a way defined by the given evolutionary rules. The success of cooperators against defectors depends on the payoffs, noises, connectivity structure, and some other details of the evolutionary rule. Now we survey those spreading mechanisms of cooperation that increase the density of cooperators and help the whole community to reach higher total income.

2. The models. The spatial versions of the evolutionary Prisoner's Dilemma games were introduced by Nowak *et al.* [14,15] and these cellular automata were capable to explain the maintenance of cooperation for the iterated games with a limited range of interaction if the players follow one of the two simplest strategies. For these strategies, denoted shortly as D and C, the player chooses always defection and cooperation, respectively. We assume that each players is located on a site x of a lattice or graph and their strategy is denoted by two-dimensional unit vectors, namely

$$\mathbf{s}_x = D = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{or} \quad C = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{1}$$

The total payoff U_x of player x comes from games with the neighbors $y \in \Omega_x$, that is,

$$U_x = \sum_{y \in \Omega_x} \mathbf{s}_x^+ \mathbf{A} \cdot \mathbf{s}_y, \tag{2}$$

where $\mathbf{s}_x^+$ denotes the transpose of the state vector $\mathbf{s}_x$, the summation runs over sites of the neighborhood Ω_x. In the present discussion the self-interaction is excluded and the rescaled payoff matrix

$$\mathbf{A} = \begin{pmatrix} 0 & b \\ c & 1 \end{pmatrix}, \quad 1 < b < 2 - c, \quad c < 0. \tag{3}$$

will be considered in the limit $c \to -0$ [14,15]. For these symmetric two-person games both interacting players receive 1 if they follow the C strategy. Conversely, each player receives 0 if both choose D. The highest payoff b is achieved by the defector against the cooperator receiving nothing. The parameter b describes the strength of temptation to choose defection. Notice, that the total payoff (b) for the defector-cooperator co-players is less than those received by two cooperators.

In the spatial evolutionary games the equivalent players play repeated games with their neighbors and they can modify their own strategy between games with neighbors. For the non-spatial models the players are located on the sites x of a graph (network) where the edges define the connected players (neighbors) and the degree z_x gives the number of neighbors for the player residing at site x.

For most of the evolutionary games the players wish to maximize their own total payoff by adopting one of the more successful strategies available in their neighborhood. This type of dynamics is analogous to the Darwinian selection in the ecological systems [16,17]. There are different ways (algorithms) defining how the players choose a new strategy. In this work our discussion will be restricted to two strategy update mechanisms, namely, synchronized update and random sequential pairwise comparison.

2.1. *Synchronized update.* This evolutionary rule was suggested and studied by Nowak *et al.* [14,15] and by many others (see e.g., [18]) during the subsequent years.

For the synchronized update the system is analogous to a cellular automaton where the strategy distribution $\{\mathbf{s}_x(t)\}$ is given in discrete time steps, i.e., $t = 0, 1, 2, \ldots$. In the subsequent time step the new strategy distribution $\{\mathbf{s}_x(t+1)\}$ is determined by the payoff distribution $U_x(t)$ that can be evaluated by using Eq. (2). According to the rule suggested by Nowak and May [14], at time $t+1$ the player x will follow the strategy of the neighboring player having the highest total payoff at time t. The generalization of this rule to arbitrary networks is straightforward and the results are discussed in papers [19-22].

2.2. *Random sequential pairwise comparison.* In this case the players refresh their strategy individually by repeating the subsequent elementary steps. We choose two neighboring players at the sites x and y at random and we evaluate their payoff (U_x and U_y) given by Eq. (2). Player x adopts the neighboring strategy $\mathbf{s}_y$ with a probability depending on the payoff difference

$$W[\mathbf{s}_x \to \mathbf{s}_y] = \frac{1}{1 + \exp\left[(U_x - U_y)/K\right]}, \tag{4}$$

where K (called temperature) characterizes the magnitude of noise [23-26]. It is worth mentioning that this evolutionary rule allows irrational choices, that is the players can adopt the less successful strategies with a low probability.

The above strategy adoption (learning) rules do not change the homogeneous states ($e.g., s_x = D$). These spatial distributions are called absorbing states and whenever one of the homogeneous states is reached the system remains there forever. During Monte Carlo simulations this phenomenon can be observed frequently for small system sizes after a transient time whose average value increases very fast with N. For large systems the inside of homogeneous regions remains unchanged because variation can occur only for those sites where one of the neighbors follows the opposite strategy.

Henceforth we assume that the system is started from a random initial state for sufficiently large sizes to avoid the mentioned small size effect. When repeating the described elementary steps the system evolves towards a state which can be characterized by the average density ρ of cooperators. Before considering the numerical results for different neighborhood structures we briefly survey the prediction of mean-field theory.

3. Prediction of mean-field theory. Within the framework of mean-field theory the system is characterized by the average density ρ of cooperators. This means that the average number of cooperators in the neighborhood of a given player is $z\rho$ if each site has z neighbors. Consequently, the average payoff for C and D strategies are

$$U_C = z\rho \quad \text{and} \quad U_D = z\rho b. \tag{5}$$

Assuming that the strategy adoption takes place between two randomly chosen players following D and C strategy then the dynamical rule with random sequential pairwise comparison [defined by Eq. (4)] yields the following equation of motion for the concen-

tration of cooperators:

$$\frac{\partial \rho}{\partial t} = \rho(1-\rho)[W(D \to C) - W(C \to D)] \tag{6}$$
$$= -\rho(1-\rho)\tanh\left(\frac{U_D - U_C}{2K}\right).$$

In this system ρ tends to zero as $t \to \infty$ since $U_D > U_C$ independently of the initial state and the value of z and K. Consequently, the cooperators die out for those evolutionary Prisoner's Dilemma games where the conditions of mean-field approximation are satisfied. Here it is worth mentioning that ρ tends to 1 for $b < 1$ representing the so-called stag-hunt game [5-8].

It is emphasized that the conditions of mean-field theory are satisfied for infinite range of interaction in a large system ($z = N-1$), or even in a spatial model where the player's distribution is well mixed before they play game with each other.

According to the mean-field theory significant reduction is expected in ρ if the spatial distribution of C and D strategies is random (uncorrelated). This phenomenon can be observed in many models after starting the system from a random initial state (see below).

4. Spreading of cooperation for synchronized strategy update. Now we consider a cellular automaton type system where the players are distributed on a square lattice and their income comes from Prisoner's Dilemma games with their first- and second neighbors ($z = 8$). The strategies are refreshed simultaneously at times $t = 1, 2, \ldots$ in such a way [14] that each player adopts the best strategy available in her neighborhood.

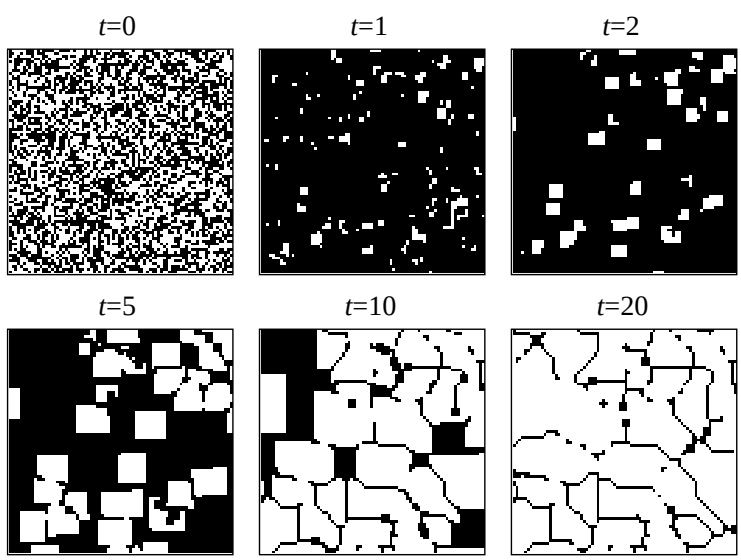


Fig. 1. Evolution of the spatial distribution of C (white) and D (black) strategies on a square lattice for a cellular automaton rule based on the adoption of the best strategies from the nearest and next-nearest neighbors in discrete time steps t if the system is started from a random initial state at $b = 1.31$.

In order to demonstrate the time evolution of the spatial distribution of strategies Figure 1 shows six consecutive snapshots on an 80×80 portion of a larger system with periodic boundary conditions. The system is started ($t = 0$) from a random distribution of D and C strategies with $\rho(t = 0) = 0.5$. In agreement with the prediction of mean-field theory a sudden reduction of ρ can be observed after the first step. In fact, only the sufficiently large rectangular blocks of cooperators could remain alive against the invasion of defectors. In the next steps the rectangular blocks of cooperators grow linearly with time because along the horizontal and vertical boundaries (separating the domains of D and C strategies) C invasions occur due to the support received by Cs from their background. Quantitatively, along the horizontal and vertical boundaries the cooperators receive $U_C = 5$ meanwhile the defectors' payoff is $U_D = 3b$. The cooperator invasion is stopped when two fronts meet, more precisely, when the regions of cooperators are separated by defectors who can exploit two cooperators at least. This is the way how the spatial distribution of strategies evolves towards a "frozen" pattern where the large regions of cooperators are separated by (one-site thick) chains of defectors. The final pattern depends on the initial state and the value of b (for details see refs. [14,15]). In general, the average density of cooperators is determined by a dynamical balance between the cooperator invasions and defector invasions. Defector invasions occur along the irregular fronts. On the contrary, cooperator invasions can be maintained step by step along the (horizontal and vertical) straight line fronts due to the simultaneous strategy refreshment. This latter process is forbidden for random sequential updates providing a less efficient way to maintain the cooperative behavior.

5. Spreading of cooperators for random sequential pairwise comparison. First we consider the spreading of cooperators on a kagome lattice ($z = 4$) when the strategy adoption mechanism is defined by the random sequential pairwise comparison described above (see subsection 2.2). One can easily check that the cooperative behavior cannot spread away from a single cooperator surrounded by defectors in the zero temperature limit ($K \to 0$). As well as for synchronized strategy updates, the adoption of cooperator strategy can become attractive if the C players form compact colonies. On the kagome lattice three cooperators forming a triangle (see Figure 2) can support the spreading of cooperation in the zero temperature limit.

Figure 2 shows the possible variations in the spatial distribution of strategies for consecutive steps if initially three (neighboring) cooperators are located in the sea of defectors. In this situation the highest income is received by the three cooperators ($U_C = 2$) meanwhile the neighboring defectors have a lower income ($U_D^{(n)} = b$) and the rest of defectors receive nothing ($U_D^{(r)} = 0$). In the zero temperature limit the only possible event is that one of the neighboring defectors adopts the cooperator strategy. The total income of this new cooperative player, however, is decreased ($U_C^{(nn)} = 1$) and becomes less than those of her neighboring defectors. Consequently, the most likely strategy change is that this player turns back to defection again and a similar temporary cooperation can occur several times in the neighborhood of the cooperator triplet. The new-born cooperator strategy can be fixed if her most successful neighboring defector

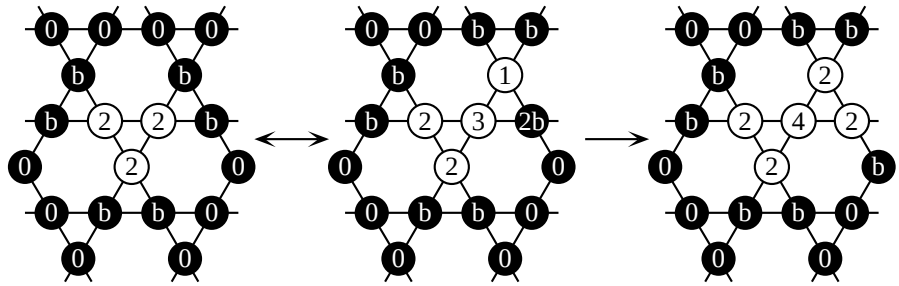


Fig. 2. Appearance of a new cooperator triplet (white bullets) in the sea of defectors (black bullets) for the random sequential pairwise comparison rule (in the limit $K \to 0$) if initially only one cooperator triplet exists on the kagome lattice. Figures in bullets refer to payoff dependent on the neighboring strategies. The arrows indicate the direction of possible transitions.

($U_D^{ms} = 2b$) has an opportunity to adopt strategy from the most successful cooperator ($U_C^{ms} = 3$). After these two consecutive steps two stable cooperator triplets (with a common player) occur and the iteration of this process yields a growing tree of cooperator triplets. The growth of the tree of triangles can be stopped by defectors separating two branches of trees formed by cooperator triplets.

For the present structure the above spreading mechanism of cooperation occurs if the temptation to choose defection does not exceed a critical value, that is, $b < b_c(K = 0) = 3/2$ in the limit $K \to 0$. For finite temperatures the efficiency of this spreading process is decreased by the irrational strategy adoptions. As a result, ρ decreases monotonically when b is increased for fixed temperature K as demonstrated in Figure 3.

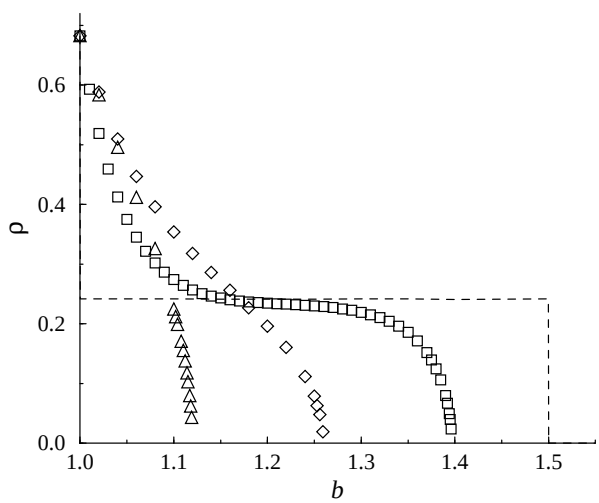


Fig. 3. Monte Carlo data for the average ρ density of cooperators versus b on the kagome lattice at three different temperatures: $K = 0.1$ (squares); 0.2 (diamonds); 0.5 (triangles). The zero temperature limit is illustrated by the dashed line.

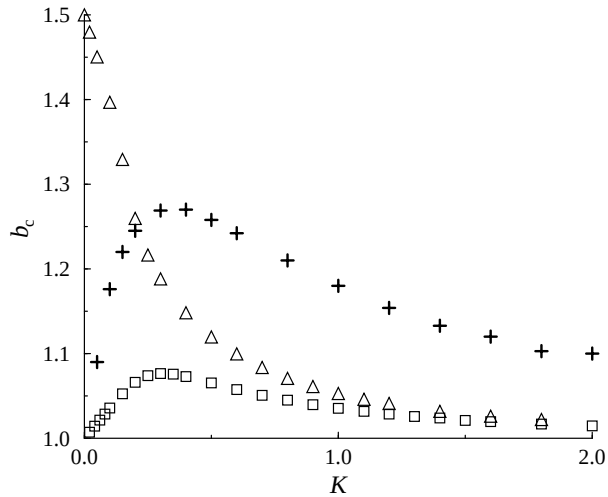


Fig. 4. The critical value of b as a function of K when using pairwise comparison for three different lattices where each site has four neighbors (triangles: kagome lattice; squares: square lattice; pluses: Bethe lattice).

The cooperative strategies can be maintained (i.e., $\rho > 0$) for $b < b_c(K)$ where $b_c(K)$ is a monotonously decreasing function of K as illustrated by triangles in Figure 4. Using Monte Carlo simulations similar behavior is found for many other d-dimensional ($d > 1$) lattices (e.g., two-dimensional square lattice with first- and second neighbor interactions ($z = 8$), triangular lattice ($z = 6$), and body centered cubic lattice ($z = 8$)) which can be spanned by overlapping triangles [27]. It is conjectured that overlapping triangles in the connectivity structure can support the spreading of cooperators although for most of the investigated structures the overlapping triangles can have two common lattice sites that makes the spreading mechanism of cooperation more complicated. Only one exception is found until now that does not satisfy this conjecture, namely, the one-dimensional lattice with first- and second neighbor interactions ($z = 4$) whose behavior is strongly affected by its one-dimensional character. The rigorous investigation of the evolutionary Prisoner's Dilemma game on this structure is in progress.

Significantly different behavior is found by Monte Carlo simulations when considering the same model on the square and Bethe lattices for $z = 4$ as illustrated in Figure 4. For both cases $b_c(K)$ has a local maximum for an optimum temperature and $b_c(K)$ tends to 1 if $K \to 0$. These structures do not support the spreading of C strategies if $K \to 0$ as demonstrated in Figure 5.

Let us consider what happens if we have four cooperators forming a square (2×2) block in the sea of defectors on the square lattice. For such an arrangement the cooperators have the highest income ($U_C = 2$), the neighboring defectors can exploit only one cooperator (i.e., $U_D^{(n1)} = b$), and all the other defectors receive nothing. Consequently, the cooperators will not adopt the defector strategy in the low temperature limit. At the same time one of the neighboring defectors can easily adopt the C strategy and this event will be so beneficial to one of the neighboring defectors (receiving $2b$) that her strategy can

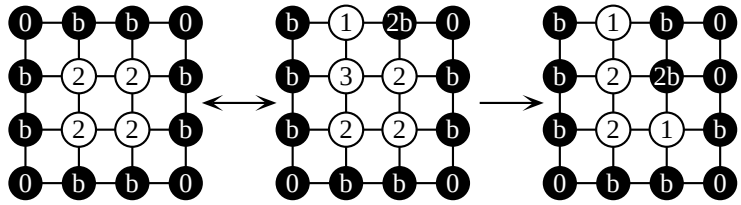


Fig. 5. Relevant consecutive configurations during the destruction of a 2×2 block of cooperators (located in the sea of defectors) when the strategy adoption is governed by random pairwise comparison on the square lattice.

be adopted by both of her neighboring C players in the subsequent strategy adoption process. Consequently, the stable block of four cooperators will be destroyed because the rest of cooperators are not capable to survive in the zero temperature limit.

The (isolated) closed loops of cooperators remain stable configurations in those structures where the neighboring defectors are capable to exploit only one cooperator if the irrational strategy adoption ($U_x > U_y$) is forbidden. The evolutionary rule defined by Eq. (4), however, allows irrational choices (if $K > 0$) which are capable to destroy these loops. After adopting a defector within the loops of cooperators the system behavior becomes analogous to a one dimensional problem with $z = 2$ where the cooperator sites are invaded by defectors with a constant average velocity. Consequently, in the absence of an effective reconstruction mechanism (e.g., a cooperation spreading process described above) the isolated loops of cooperators vanish exponentially with a typical relaxation time proportional to $\exp[(2-b)/K]$. This feature causes serious difficulties when considering these systems by Monte Carlo simulations for sufficiently low temperatures.

The appearance of cooperators through noise effects is supported for those constellations where the C strategies form colonies because within these colonies the cooperators help each other. In general, the probability of a $C-C$ pair is higher than ρ^2 predicted by the mean-field theory. This fact is due to the present evolutionary rule favoring the appearance of $C-C$ and $D-D$ pairs that gives an extra advantage to cooperators against defectors. For large temperatures, however, this type of correlation is reduced and the system behavior can be well approximated by a mean-field theory predicting $b_c = 1$. Figure 4 shows that for the spatial models $b_c(K) \to 1$ if K goes to infinity. Here we have to emphasize that the high temperature limit is analogous to the so-called weak selection limit used in the biological literature, e.g., [28-30].

In agreement with the expectations the mean-field type behavior is enforced when increasing the number of neighbors independently of the evolutionary rules [14,15,31-35]. In order to check this tendency for the pairwise comparison, the above investigations were repeated on many other lattice structures characterized by $z = 4$, 6 and 8. According to the preliminary numerical results the highest critical value of b_c (namely, $b_c(K=0) = 3/2$) is found on the kagome lattice in the limit $K \to 0$.

The existence of an optimum temperature in the maintenance of cooperation is accompanied by a proper temperature-dependence in the density of cooperators for a fixed

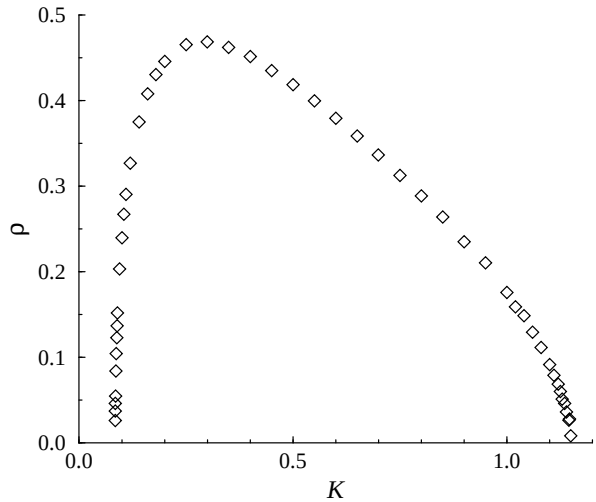


Fig. 6. Average density of cooperators as a function of temperature K on the square lattice for $b = 1.03$.

b value as shown in Figure 6 [26]. This behavior resembles the coherence resonance [36-38] that was already predicted for some evolutionary games [39] and many other noise driven excitable systems [40].

The extinction process of cooperators at the edges of coexistence region ($K = K_{c1}$ and K_{c2}) exhibits the features of a critical transition belonging to the directed percolation universality class [25,41]. In other words, ρ vanishes as

$$\rho \simeq |K - K_c|^{\beta}. \tag{7}$$

where $K_c = K_{c1}(b)$ or $K_{c2}(b)$ and $\beta = 0.58(1)$ for the two-dimensional systems. The algebraic behavior of ρ is accompanied with a power law divergency in the density fluctuation, correlation length, and relaxation time. This universal behavior is related to the existence of an absorbing state (for details see [42]). As mentioned above the present system has two homogeneous absorbing states, therefore the extinction of defectors (appearing for suitable values of parameters) has similar features [6,25,41].

The relevance of the above mentioned topological features is confirmed by the generalized versions of mean-field (or pair) approximations when we determine the probability of each possible configuration on a set of small clusters formed by n lattice sites (for technical details see [6]). Evidently, this method becomes more and more reliable if we choose a larger n. As mentioned above, when we increase the value of b the traditional mean-field theory ($n = 1$) predicts a sudden change from the homogeneous D state to the homogeneous C state at $b = b_c^{mf}(K) = 1$ independently of the temperature. The pair approximation ($n = 2$) is already capable to describe qualitatively the coexistence of C and D strategies, however, the numerical results show poor agreement with the Monte Carlo data, particularly at the zero temperature limit where its prediction becomes qualitatively wrong ($b_c^{(pa)}(K) \to 2$ if $K \to 0$). The three- and five-site approximations on

the kagome lattice and also the four- and nine-site approximations on the square lattices could reproduce the Monte Carlo data with an adequate accuracy [25] because these clusters of sites take into account the loops in the connectivity structures.

There is another way to demonstrate the importance of small (entangled) loops in the connectivity structures. Namely, the evolutionary prisoner's dilemma game can also be studied on the (loop-free) Bethe lattice with using the generalized versions of mean-field theories. Similar Monte Carlo results are expected on a sufficiently large random regular graph that becomes locally similar to the Bethe lattice because the typical (minimal) loop size is proportional to $\ln N$ [42]. In Figure 4 the plus symbols represent the Monte Carlo data obtained on a random regular graph of degree $z = 4$ if we use the same evolutionary rules for $N = 10^6$. To reproduce this behavior with the generalized mean-field theory we had to extend this analysis to the level of eleven-site clusters. The K-dependence of the critical value of temptation ($b_c^{(rrg)}(K)$) on the random regular graph is qualitatively similar to those found on the square lattice, although $b_c^{(rrg)}(K)$ is significantly larger than $b_c^{(sl)}(K)$ for arbitrary temperatures. The high temperature (weak selection) behavior of a similar evolutionary prisoner's dilemma game was analyzed by Ohtsuki *et al.* [28] on finite graphs, and their results allow the coexistence of C and D strategies within a range of b (if $K \to \infty$). It is hoped that the asymptotic behavior in these types of models will be clarified within a few years.

One of the most relevant messages of the numerical data plotted in Figure 9 is that the kagome lattice provides the best lattice connectivity structure to achieve the highest critical b_c value for low temperatures. For the opposite case ($K \to \infty$), the loop-free structures can help the maintenance of cooperative behavior most efficiently if the number of neighbors (degree) is fixed (here $z = 4$). In the subsequent section we will consider the effect of inhomogeneous degree distribution on the density of cooperators if the connectivity structure is defined by the so-called scale-free networks.

6. Enhancement of cooperation on scale-free networks. In this section we discuss briefly the results of Santos *et al.* [44,45] who considered two-strategy prisoner's dilemma games with several evolutionary rules on a random network suggested by Barabási and Albert [46]. One of the applied strategy adoption rule was similar to the mentioned pairwise comparison at a low temperature. It is reported that the average density of cooperators was surprisingly large ($\rho > 0.8$) in the whole region of temptation ($1 < b < 2$) and the relevant results are not affected significantly by the details of evolutionary rules.

The creation of the Barabási-Albert random network can start from three connected sites and it is grown by adding a new site to the system in each elementary step. The new sites are linked to two existing sites chosen randomly with probabilities proportional to the actual degree (further details in [46]). After $(N - 3)$ elementary steps the resultant graph consist of N sites which can have different number of neighbors (z) in such a way that the probability to find z neighbors at a given sites is proportional to $f(z) \sim 1/z^3$ if z does not exceed a natural limit dependent on N. During the growth process the average number of neighbors tends to $\langle z \rangle = 4$. This connectivity structure contains sites (called hubs) that has a large number of neighbors. In general, the large hubs are created at

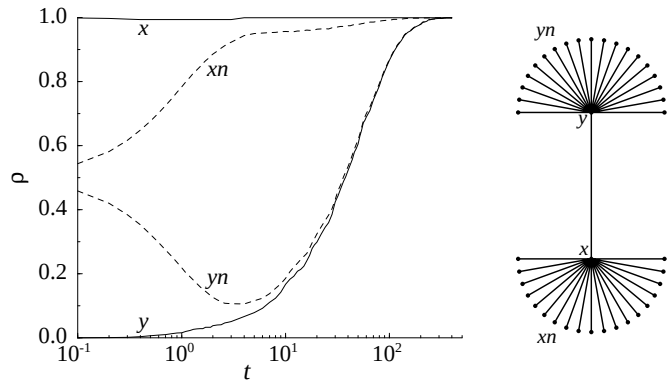


Fig. 7. Time-dependence of the frequency of cooperators on the sites of cooperator (x) and defector (y) hubs and on their neighborhood (xn and yn). The plotted data are averaged over 1000 runs for $R = 0.5$ characterizing the effects of neglected links and noise.

the beginning of network creation and these sites are typically linked to other hub(s). This latter feature plays a crucial role in the mechanism supporting the spreading of cooperation as described by Santos *et al.* [44,45]. Now we will discuss this mechanism throughout a numerical example.

Let us consider a small portion of players residing on two connected hubs and their neighboring sites as illustrated schematically in the right hand plot of Figure 7. For sake of simplicity we assume that the connected hubs (denoted by x and y) have $N_{xn} = N_{yn} = 49$ neighbors whose further links are not denoted. Thus the whole subsystem has $N = 2 + N_{xn} + N_{yn} = 100$ sites. It is also assumed that the number of neighbors at the neighboring sites (xn and yn) is significantly less than 50. To demonstrate the competition between the cooperator and defector hubs the system is started from an initial state where $\mathbf{s}_x(t = 0) = C$ and $\mathbf{s}_y(t = 0) = D$, and their neighborhood follows the D and C strategies with the same probability, that is, $\rho_{xn}(t = 0) = \rho_{yn}(t = 0) = 0.5$. In the model suggested by Santos *et al.* the evolutionary rule is controlled by the difference of total payoffs favoring hubs. Consequently, the strategy adoption process can be simplified in the subsystem.

During the strategy adoption process on this subsystem we choose randomly a player to adopt strategy from one of her neighbors selected also at random. If one of the hubs (x or y) has a chance to adopt strategy from the neighbors of low degree (xn or yn) then nothing happens because generally $U_x >> U_{xn}$ and $U_y >> U_{yn}$. In the opposite case, the xn and yn players adopt the current strategy of players x and y, respectively. The effect of the neglected interactions and learning connections are taken into consideration by allowing the players xn and yn to adopt a strategy with a probability R from a player selected randomly in the whole subsystem. The results of numerical investigations are shown in Figure 7.

Figure 7 illustrates what happens if the connected hubs follow opposite strategies at the beginning. Initially, the total income of the cooperator hub is lower then those received by the defector hub because they practically have an equal number of coopera-

tive co-players. However, each hub transfers her strategy to her neighborhood and these changes in their neighborhood are beneficial to cooperator hubs and disadvantageous to defector hubs. Consequently, after a short transient time the cooperator hubs become the best players whose strategy is adopted by all the other (linked) players. This phenomenon is robust if the degree of hubs exceeds significantly the maximum number of neighbors for their co-players. The robustness of this mechanism can be confirmed by the present numerical example within a wide range of parameters (N_x, N_y, R and b). Further systematic investigations are required to clarify the effect of noise (temperature) and degree distribution on the average density of cooperators in the large N limit.

In the original models considered by Santos *et al.* the simulations were performed for different values of N. For small sizes the result of simulation depends on the graph and initial strategy distribution. As a result, the state of the isolated defector hub(s) can remain practically unchanged during the simulations. A similar phenomenon is reported by Gómez-Gardañes *et al.* [47] who have considered the same model with a synchronized strategy update. Santos *et al.* [44,45] have checked the consequence of the removal of links connecting large hubs. In agreement with the expectation, the numerical investigation indicated a reduction in the average density of cooperators.

The additional effect of cooperation spreading mediated by the overlapping triangles on the density of cooperators was also investigated by Santos *et al.* when considering the model on another scale-free network suggested by Dorogovtsev *et al.* [48]. In this case the network creation is slightly modified: the additional new sites are linked to both ends of an edge selected at random. This creation procedure yields a network that can be spanned by (one-site) overlapping triangles meanwhile the degree distribution remains practically unchanged. On this network the simulations indicated a slight increase in ρ in comparison to results obtained on the Barabási-Albert network.

The topological conditions of the mentioned mechanism are well satisfied for the scale-free networks as well as for many other connectivity structure exhibiting connected large hubs. There exist, however, many other non-regular (social) networks where the inhomogeneity in the degree distribution is not so relevant (e.g., Erdős-Rényi graphs [49] and Watts-Strogatz small-world structure [50]). The numerical investigations of the evolutionary prisoner's dilemma games on different non-regular structures have demonstrated some improvement in the cooperative behavior related to the inhomogeneity in the number of neighbors [21,51-55].

The above models on non-regular networks involve intrinsically an inhomogeneous strategy adoption probability due to the entanglement between connectivity structure and dynamical rules [52,53]. Namely, the strategy adoption becomes asymmetric between two neighbors if they have a different number of neighbors when using the above described evolutionary rules. In the following section we discuss particular cases where the spatial network is regular while the dynamical rule is inhomogeneous.

7. Enhancement of cooperation for inhomogeneous dynamics. Remarkable increase of cooperation is observed for several systems where the inhomogeneous imitation activity is built into the models artificially to characterize the asymmetric and different

influence of players to each other [51,55,56]. For example, Wu *et al.* [55,56] have considered a model where the influence of player y on her neighbor x is quantified by a random parameter w_{xy} $(0 < w_{xy} \neq w_{yx} < 1)$ affecting the preferential selection of a neighbor whom the strategy can be adopted from.

The numerical justification of the enhancement of cooperation will be demonstrated for one of the simplified versions of the model [57] suggested originally by Wu *et al.* [55,56]. For sake of simplicity we assume that w_{xy} is described by one of the two possible values in such a way that the spatial distribution w_{xy} is quenched during the simulations. Now our investigation will be focused on the spatial distribution of w_{xy} that provides the highest enhancement of cooperation.

In the present model we have two types of players (denoted as A and B) distributed randomly on a two-dimensional lattice before the start of simulation. The fixed spatial distribution of players is defined by a set of site variables $n_x = A$ or B and characterized by densities ν and $(1 - \nu)$. Independently of n_x the players can follow one of the $s_x = C$ or D strategies and their total payoff U_x comes from one Prisoner's Dilemma game with the neighbors as given by the expression (2).

The type of players becomes relevant in the iterated strategy adoption processes controlling the variation of strategy distribution. More precisely, the pairwise comparison evolutionary process (detailed in Section 2.2) is modified:

$$W(s_x \to s_y) = w_{xy} \frac{1}{1 + \exp[(U_x - U_y)/K]} \tag{8}$$

where temperature K characterizes the uncertainties (noises) and the pre-factor w_{xy} is given as

$$w_{xy} = \begin{cases} 1, & \text{if } n_y = A, \\ w, & \text{if } n_y = B, \end{cases} \tag{9}$$

and the fixed value of w $(0 < w < 1)$ describes the strength of reduced teaching activity if the site y is occupied by a player of type B. This means that here the players A and B have different teaching activity (or efficiency). Such a feature can occur in some communities where the young and old (or attractive and repulsive) persons have different capability to transfer their own strategy.

Figure 8 summarizes the results of Monte Carlo simulations when varying the density of B players on a kagome lattice for fixed b and K values. In this system the stationary states obtained for $\nu = 0$ and 1 are equivalent because the homogeneous reduction of strategy adoptions yields only slower evolution towards the same stationary state. The largest increase in ρ can be observed if the portion of A and B types players are approximately equivalent. Notice that the increase of the density of cooperators $(\Delta\rho)$ varies linearly with the concentration of the minority type of players, i.e., $\Delta\rho \simeq \alpha\nu$ if $\nu << 1$ and $\Delta\rho \simeq \alpha(1 - \nu)$ if $(1 - \nu) << 1$ where α increases monotonously with $(1 - w)$. The increase of the density of cooperators may be so large that defectors can die out within a range of ν as demonstrated in Figure 8 by data for $w = 0.1$.

Unfortunately the clear and simple explanation of this mechanism is not yet known although similar behaviors are found on the square lattice, too. For both two-dimensional

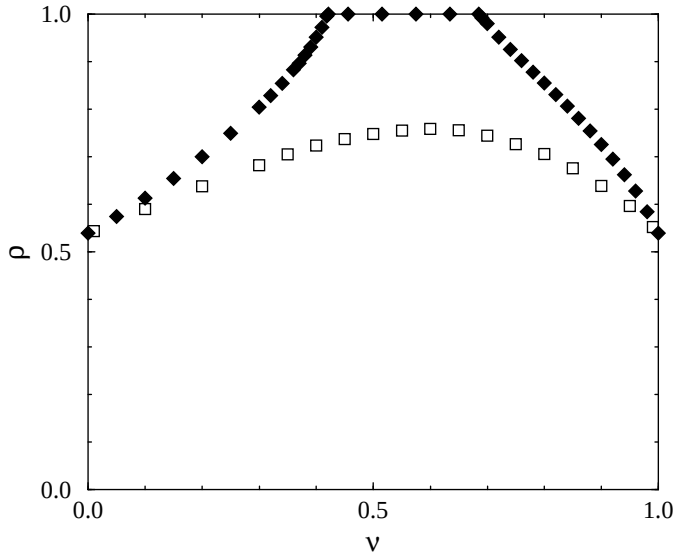


Fig. 8. Density ρ of cooperators as a function of ν (density of players with reduced teaching activity) for $w = 0.1$ and 0.2 on the kagome lattice at $b = 1.03$ and $K = 0.5$.

structures the improvement vanishes in the low temperature limit and remains remarkable even if $K \to \infty$. The main features of this phenomenon are reproduced qualitatively well by the pair approximation [57].

The hidden curiosity of this mechanism becomes striking when repeating the same analysis with choosing other possibilities in the randomness of w_{xy}. For example, significantly smaller variations of ρ are observed for those cases when the learning (instead of teaching) activity is blocked at the B type players by defining the pre-factor w_{xy} in Eq. (8) as

$$w_{xy} = \begin{cases} 1, & \text{if } n_x = A, \\ w, & \text{if } n_x = B. \end{cases} \tag{10}$$

It is worth mentioning that the spatial distribution of w_{xy} becomes equivalent for the cases defined by Eqs. (9) and (10) if the (periodical) spatial arrangement of players A and B is resembling the black and white squares on the chessboard. Interestingly, this type of dynamical inhomogeneity yields only a slight change in ρ that may be either positive or negative depending on the model parameters.

From the mentioned numerical results one can conclude that the random distribution of influential players representing asymmetric teaching-learning activity in the community can also support the maintenance of cooperative behavior [52,55-57]. In the following section we will compare the strength of all the above mentioned mechanisms.

8. Summary. In the present work we have surveyed the spreading mechanisms of cooperation supporting the cooperative behaviors in the evolutionary prisoner's dilemma games. The discussion is restricted to systems where the players can use only two strate-

gies (C and D) and the interaction between them is described by a fixed network (graph).

On the square lattice it is shown that along the horizontal and vertical fronts (separating C and D domains) the cooperator invasion is supported by their background until the income of exploiting defectors do not exceed a threshold value dependent on the neighborhood. The advance of C players along these straight line fronts can be directly utilized by synchronized strategy update rules that maintain the invasion with a constant velocity. These type of C invasion can be stopped when the front lines collide, more precisely, when the C domains are separated from each other by one-site thick layers of defectors.

The noisy rules make the straight line fronts irregular that can prevent the expansion of C domains. Despite it, the C invasion can occur for some regular connectivity structures even if the strategy adoption is controlled by random sequential pairwise comparison in the zero temperature limit. It is demonstrated that the one-site overlapping triangles in the connectivity structure (for $z = 4$) help the spreading of cooperation. In this case one can observe growing trees of cooperator triplets whose growth can be stopped by defectors separating two branches of the cooperator trees. In some sense this phenomenon is similar to those observed for the cellular automaton type rules and its efficiency is weaker as demonstrated in Figure 9 where the b-dependence of ρ is compared for different connectivity structures and evolutionary rules.

The numerical and analytical investigations of the systems with random sequential update indicate that the cooperation vanishes in the zero temperature limit for those connectivity structures which can not be spanned by overlapping triangles. In these latter cases, however, there exists an optimum temperature (noise level) where cooperators can exist with a sufficiently high density as illustrated in Figure 9. This means that the irrational strategy adoptions can also help the survival of cooperators forming colonies. It is found, furthermore, that the loop-free (non-spatial) connectivity structures provide the best regular connectivity structure for the cooperators to exist in the high temperature limit.

For the non-regular connectivity structures the connected large hubs support the emergence of cooperative behavior. The corresponding mechanism is robust on the scale-free graphs and capable to maintain cooperation with a high density ($\rho > 0.8$) in the whole region of payoff b (see Figure 9). In the models suggested by Santos *et al.* the enhanced role of large hubs comes from the fact that the players compare their total income increasing with the number of neighbors. It is found that the efficiency of this mechanism is reduced significantly if the strategy adoption rule is based on the normalized payoff difference [e.g., $(U_x/z_x - U_y/z_y)$ where z_x is the number of neighbors at the site x]. For the non-regular connectivity structures the players are not equivalent and the symmetry between the neighboring players is broken for most of the evolutionary rules we can apply. It turns out that the asymmetry in the master-follower role between two neighboring players can also result in an improvement in the cooperative behavior.

Further systematic investigations of the discussed phenomena are in progress. The current studies can explore the interference between the mentioned mechanisms. We have

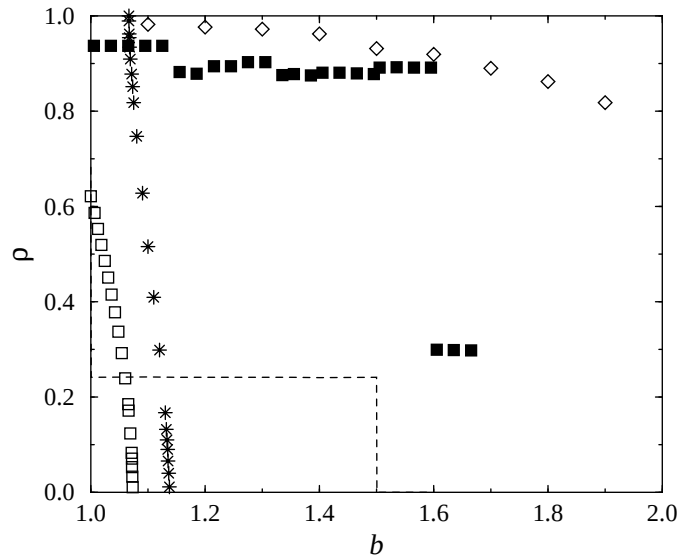


Fig. 9. Comparison of the density of cooperators as a function of b for different connectivity structures. The dashed line shows the results on the kagome lattice at the zero temperature limit. Open squares represent data for the square lattice with pair-wise comparison at the optimum temperature ($K = 0.4$) [26]. Diamonds denote Monte Carlo data obtained by Santos et al. [44,45] on the scale-free Barabási-Albert network. Closed squares refer to data on the square lattice with first- and second-neighbor interactions when using synchronized strategy update suggested by Nowak and May [14]. Stars refer to evolutionary game with a modified pairwise comparison where the teaching activity is blocked by a factor of $w = 0.1$ for half of the players distributed randomly on a square lattice at $K = 1.0$ [57].

to emphasize that recently the research of evolutionary games expands progressively. In the last years further promising results were reported including the consideration of the multilevel selection [58,59], the co-evolution of strategy distribution and connectivity structure [60,61], and the separation of learning and interaction graphs [62,63].

References

[1] M. A. Nowak, *Evolutionary dynamics of biological games*, Science 303 (2004), 793–799.

[2] K. Lindgren and G. Nordahl, *Evolutionary dynamics of spatial games*, Physica D 75 (1994), 292–309.

[3] M. Nakamaru, H. Masuda and Y. Iwasa, *The evolution of cooperation in a lattice-structued population*, J. Theoretical Biology 184 (1997), 65–81.

[4] J. Miękisz, *Statistical mechanics of spatial evolutionary games*, J. Physics A: Math. Gen. 37 (2004), 9891–9906.

[5] J. Miękisz, *Evolutionary game theory and population dynamics*, arXiv:q-bio/0703062 (2007).

[6] G. Szabó and G. Fáth, *Evolutionary games on graphs*, Physics Reports (2007), in press.

[7] H. Gintis, *Game Theory Evolving*, Princeton University Press, Princeton, 2000.

[8] R. Cressman, *Evolutionary Dynamics and Extensive Form Games*, MIT Press, Cambridge MA, 2003.
[9] M. A. Nowak and K. Sigmund, *Arithmetics of mutual help*, Scientific American 272 (1995), 76–81.
[10] W. Hamilton, *Genetical evolution of social behavior II*, J. Theoretical Biology 7 (1964), 17–52.
[11] R. Axelrod, *The Evolution of Cooperation*, Basic Books, New York, 1984.
[12] C. Hauert, S. De Monte and K. Sigmund, *Volunteering as Red Queen mechanism for cooperation in public goods game*, Science 296 (2002), 1129–1132.
[13] M. A. Nowak, *Five rules of the evolution of cooperation*, Science 314 (2006), 1560–1563.
[14] M. A. Nowak and R. M. May, *The spatial dilemmas of cooperation*, Int. J. Bifurcation and Chaos 3 (1993), 35–78.
[15] M. A. Nowak, S. Bonhoeffer and R. M. May, *More spatial games*, Int. J. Bifurcation and Chaos 4 (1994), 23–56.
[16] R. Axelrod and W. D. Hamilton, *The evolution of cooperation*, Science 211 (1981), 1390–1396.
[17] J. Maynard Smith, *Evolution and the Theory of Games*, Cambridge University Press, Cambridge, 1982.
[18] T. Killingback and M. Doebeli, *Self-organized criticality in spatial evolutionary game theory*, J. Theoretical Biology 191 (1998), 335–340.
[19] G. Abramson and M. Kuperman, *Social games on social network*, Physical Review E 63 (2001), 030901(R).
[20] N. Masuda and K. Aihara, *Spatial prisoner's dilemma optimally played in small-world networks*, Physics Letters A 313 (2003), 55–61.
[21] O. Duran and R. Mulet, *Evolutionary prisoner's dilemma in random graphs*, Physica D 208 (2005), 257–265.
[22] J. Gómez-Gardeñes *et al.*, *Dynamical organization of cooperation in complex topologies*, Physical Review Letters 98 (2007), 108103.
[23] L. E. Blume, *The statistical mechanics of best-response strategy revision*, Games and Economic Behavior 11 (1995), 111–145.
[24] L. E. Blume, *How noise matters*, Games and Economic Behavior 44 (2003), 251–271.
[25] G. Szabó and C. Tőke, *Evolutionary prisoner's dilemma game on a square lattice*, Physical Review E 58 (1998), 69–73.
[26] G. Szabó, J. Vukov and A. Szolnoki, *Phase diagrams for an evolutionary prisoner's dilemma game on two-dimensional lattices*, Physical Review E 72 (2005), 047107/1–4.
[27] J. Vukov, G. Szabó and A. Szolnoki, *Cooperation in the noisy case: Prisoner's dilemma game on two types regular random graphs*, Physical Review E 74 (2006), 067103/1–4.
[28] H. Ohtsuki, C. Hauert, E. Lieberman and M. A. Nowak, *A simple rule for the evolution of cooperation on graphs and social networks*, Nature 441 (2006), 502–505.
[29] C. Taylor, Y. Iwasa and M. A. Nowak, *Symetry of fixation times in evolutionary dynamics*, J. Theoretical Biology 243 (2006), 245–251.
[30] A. Traulsen, J. M. Pacheco and M. A. Nowak, *Pairwise comparison and selection temperature in evolutionary game dynamics*, J. Theoretical Biology 246 (2007), 522–529.
[31] M. Ifty, T. Killingback and M. Doebeli, *Effect of neighborhood size and connectivity on the spatial continuous prisoner's dilemma*, J. Theoretical Biology 231 (2004), 97–106.
[32] F. C. Santos, J. F. Rodrigues and J. M. Pacheco, *Graph topology plays a determinant role in the evolution of cooperation*, Proc. Royal Society London B 273 (2006), 51–55.

[33] C.-L. Tang, W.-X. Wang, X. Wu and B.-H. Wang, *Effects of average degree on cooperation in networked evolutionary game*, Eur. Phys. J. B 53 (2006), 411–415.

[34] Z.-X. Wu and Y.-H. Wang, *Cooperation enhanced by the difference between interaction and learning neighborhoods for evolutionary spatial prisoner's dilemma games*, Physical Review E 75 (2007), 041114/1-7.

[35] M. H. Vainstein, A. T. C. Silva and J. J. Jeferson, *Does mobility decrease cooperation?*, J. Theoretical Biology 244 (2007), 722–728.

[36] M. Perc, *Coherence resonance in a spatial prisoner's dilemma game*, New J. Physics 8 (2006), 22.

[37] M. Perc and M. Marhl, *Evolutionary and dynamical coherence resonances in the pair approximated prisoner's dilemma game*, New J. Physics 8 (2006), 142.

[38] M. Perc, *Double resonance in cooperation induced by noise and network variation for an evolutionary prisoner's dilemma*, New J. Physics 8 (2006), 183.

[39] A. Traulsen, T. Röhl and H. G. Schuster, *Stochastic gain in population dynamics*, Physical Review Letters 93 (2004), 028701.

[40] A. S. Pikovsky and J. Kurths, *Coherence resonance in a noise-driven excitable system*, Physical Review Letters 78 (1997), 775–778.

[41] J. N. R. Chiappin and M. J. de Oliveira, *Emergence of cooperation among interacting individuals*, Physical Review E 59 (1999), 6419.6421.

[42] J. Marro and R. Dickman, *Nonequilibrium Phase Transitions in Lattice Models*, Cambridge University Press, Cambridge, 1985.

[43] B. Bollobás, *Random graphs*, Academic Press, New York, 1985.

[44] F. C. Santos and J. M. Pacheco, *Scale-free networks provide a unifying framework for the emergence of cooperation*, Physical Review Letters 95 (2005), 098104.

[45] F. C. Santos, J. F. Rodrigues and J. M. Pacheco, *Graph topology plays a determinant role in the evolution of cooperation*, Proc. R. Soc. London B 273 (2006), 50–55.

[46] A.-L. Barabási and R. Albert, *Emergence of scaling in random networks*, Science 286 (1999), 509–512.

[47] J. Gómez-Gardañes, M. Campillo, L. M. Floria and Y. Moreno, *Dynamical organization of cooperation in complex topologies*, Physical Review Letters 98 (2007), 108103.

[48] S. N. Dorogovtsev, J. F. F. Mendes and A. N. Samukhin, *Size-dependent degree distribution of a scale-free growing network*, Physical Review E 63 (2001), 052101.

[49] P. Erdős and A. Rényi, *On random graphs*, Publ. Math. Debrecen 6 (1959), 290–297.

[50] D. J. Watts and S. H. Strogatz, *Collective dynamics of 'small world' networks*, Nature 393 (1998), 440-442.

[51] M. H. Vainstein and J. J. Arenzon, *Disordered environments in spatial games*, Physical Review E 64 (2001), 051905.

[52] B. J. Kim, A. Trusina, P. Holme, P. Minnhagen, J. S. Chung and M. Y. Choi, *Dynamic instabilities induced by asymmetric influence: Prisoner's dilemma game in small-world networks*, Physical Review E 66 (2002), 021907.

[53] P. Holme et al., *Prisoner's dilemma in real-world acquintance networks: Spikes and quasiequilibria induced by the interplay between structure and dynamics*, Physical Review E 68 (2003), 030901.

[54] G. Abramson and M. Kuperman, *Social games in a social network*, Physica D 208 (2005), 257–265.

[55] Z.-X. Wu, X.-J. Xu and Y.-H. Wang, *Prisoner's dilemma game with heterogeneous influental effect on regular small-world networks*, Chinese Physical Letters 23 (2006), 531–534.

[56] Z.-X. Wu et al., *Evolutionary prisoner's dilemma game with dynamic preferential selection*, Physical Review E 74 (2006), 021107.

[57] A. Szolnoki and G. Szabó, *Cooperation enhanced by inhomogeneous activity of teaching for evolutionary prisoner's dilemma games*, Europhysics Letters 77 (2007), 30004.

[58] A. Traulsen and M. A. Nowak, *Evolution of cooperation by multilevel selection*, Proc. Natl. Acad. Sci 103 (2006), 10952–10955.

[59] J. A. Fletcher and M. Zwick, *The evolution of altruism: Game theory in multilevel selection and inclusive fitness*, J. Theoretical Biology 245 (2007), 26–36.

[60] M. G. Zimmermann and V. M. Eguíluz, *Cooperation, social networks, and the emergence of leadership in a prisoner's dilemma with adaptive local ineractions*, Physical Review E 72 (2005), 056118.

[61] J. M. Pacheco, A. Traulsen and M. A. Nowak, *Coevolution of strategy and complex networks with dynamical linking*, Physical Review Letters 97 (2006), 258103.

[62] H. Ohtsuki, M. A. Nowak and J. M. Pacheco, *Breaking the symmetry between interaction and replacement in evolutionary dynamics on graphs*, Physical Review Letters 98 (2007), 108106.

[63] Z.-X. Wu and Y.-H. Wang, *Cooperation enhanced by difference between interaction and learning neighborhoods for evolutionary spatial prisoner's dilemma games*, Physical Review E 75 (2007), 041114.

STOCHASTIC MODELS IN BIOLOGICAL SCIENCES
BANACH CENTER PUBLICATIONS, VOLUME 80
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WARSZAWA 2008

PROBLEMS WITH CLASSICAL MODELS OF SEX-RATIO EVOLUTION

KRZYSZTOF ARGASIŃSKI

Institute of Environmental Sciences, Jagiellonian University
Gronostajowa 7, 30-387 Kraków, Poland
E-mail: argas1@wp.pl

Abstract. The classical theory of the sex-ratio evolution, known as the sex-ratio game, is based on the maximization of the number of grandchildren, treated as a fitness measure of a female producing offspring of the sex ratio that is coded in her genes. The theory predicts that it is more profitable to produce offspring with less numerous sex. We can find in the literature mutually exclusive conclusions based on this prediction: some textbooks say that populations with the equal number of sons and daughters are evolutionarily stable, others identify this ratio as a stable state of a population with different individual strategies being allowed. It is also not clear whether a primary or secondary sex ratio is a target of evolution. Moreover, the classical theory ignores the role of males, who host non-expressed sex-ratio genes. Our new approach, based on multipopulation dynamic evolutionary games, shows that in populations of players with individual strategies, the secondary sex ratio is attracted by the current value of the primary sex ratio, which slowly converges to the unique stable value of 0.5. Male hosts of the sex-ratio gene are important because perturbations of a stable state of males subpopulation can destabilize the whole system.

1. Introduction. In nature, the observed ratio of males to females is in most cases 0.5. However, a single male individual can fertilize many females. In many species, dominant males can monopolize all females in the population, thus making reproduction of other males impossible. Nevertheless, also in this case the 0.5 sex ratio can be observed. The question is: what mechanism is responsible for what seems to be a suboptimal sex ratio? When males, unlike females, do not produce offspring, half of individuals are excluded from reproduction. Such a population has a lower mean fitness than a population in which majority of individuals reproduce. The question arises: what is the origin of the

2000 *Mathematics Subject Classification*: Primary 92D15.

Key words and phrases: evolutionary game theory, replicator dynamics, sex-ratio evolution, multipopulation evolutionary games.

The paper is in final form and no version of it will be published elsewhere.

 © Instytut Matematyczny PAN, 2008

0.5 sex ratio? This problem was considered by Darwin. Unfortunately, he has not solved it. First attempts to explain this phenomenon were made in the end of the 19-th century. They were also the first applications of mathematical methods in evolutionary biology. Their author was a German biologist, Carl Dusing. Using the mathematical reasoning, he argued that the natural selection could lead to the stabilization of the sex ratio on the level of 0.5. His results are founded on one basic assumption - the number of grandchildren is assumed as a fitness measure of a female who produces offspring of a sex ratio that is coded in her genes. It is known that the sex of an offspring is an effect of both a mother and a father. But a female can express her individual strategy by spontaneous abortions of foetuses with a wrong sex. This assumption has become the cornerstone of sex-ratio models. Ronald A. Fisher used a similar argument in his book and the concept is often attributed to him (see Fisher [4], Queller [9]). The Dusing-Fisher approach produces the following predictions. In the case of a shift of the population sex ratio to males, a female who produces proportionally more female offspring than other females will have a higher number of grandchildren. Similarly, when the population sex ratio is shifted to females, the production of a higher proportion of males will ensure a higher number of grandchildren. Formalization of this reasoning can be obtained by the construction of a fitness function whose arguments are: an individual strategy (that is the sex ratio in a female brood) and the population sex ratio. It can be done in the following way.

Let us take a game-theoretic perspective. Players are females, and the proportion of males in the brood of a single female, described by P_{ind} is her individual strategy. Let P denote the population sex ratio (interpreted as the proportion of males), N the number of all individuals in the population, and x the number of children of a single female. Now we can derive a payoff function which describes the number of grandchildren of a female with a strategy P_{ind}. A female with the strategy P_{ind} will have $P_{ind}x$ sons, and these sons will produce $P_{ind}x^2\frac{(1-P)N}{PN}$ grandchildren. This female will also have $(1-P_{ind})x$ daughters, who will produce $x^2(1-P_{ind})$ grandchildren. Hence the total number of grandchildren is given by the following expression:

$$\begin{aligned} P_{ind}x^2((1-P)N/PN) + x^2(1-P_{ind}) &= x^2\left[\frac{P_{ind}(1-P)}{P} + (1-P_{ind})\right] \\ &= x^2(1-P)\left[\frac{P_{ind}}{P} + \frac{1-P_{ind}}{1-P}\right]. \end{aligned} \tag{1}$$

Let us assume for simplicity that $x = 1$, then the right-hand side of (1) can be written as $(1-P)F(P_{ind}, P)$, where $F(P_{ind}, P) = \frac{P_{ind}}{P} + \frac{1-P_{ind}}{1-P}$.

As we can see, $F(P_{ind}, P)$ can be used to compare numbers of grandchildren of females with different individual strategies. This function is a payoff function of an evolutionary game, which in the literature is called the SEX-RATIO GAME (see [2], [3], [7], [8], [9], [11]). In a situation when $P = 0.5$, all strategies will obtain equal payoffs, hence this is an equilibrium. When this equilibrium is disturbed to the advantage of one of the sexes, it is more profitable to produce more individuals of the less numerous sex. In the biological literature we can find different conclusions based on this mathematical reasoning, sometimes mutually exclusive. Generally, in textbooks there exist two main variants:

- Natural selection promotes females with an individual strategy $P_{ind} = 0.5$ (see [5]).
- Natural selection leads to the 0.5 population sex ratio without favoring any individual strategy. Population is non-homogenous, there are different individual strategies (see [10]).

In biology, there is a very important distinction between the sex ratio among zygotes (called the primary sex ratio) and the sex ratio among adult individuals (called the secondary sex ratio). In textbooks, two types of relations between the primary and secondary sex ratio can be found:

- Primary sex ratios converge to 0.5 and there are no compensations of the secondary sex ratio perturbations (see [10]).
- The primary sex ratio can be biased away from 0.5, but the secondary sex ratio converges to 0.5 (see [6]).

An advantage of the Dusing-Fisher approach is its simplicity. As we have seen above, the derivation of this model takes only a few lines of text. Still, there are some questions. Does an arbitrary assumption that the number of grandchildren is a measure of fitness allows to extrapolate a reproductive success to future generations? If a female $\hat{P}_{ind}$ has more grandchildren than a female P_{ind}, then in the following generations, will $\hat{P}_{ind}$ also have more descendants than P_{ind}? Let us recall that the classical model depends on two variables: the individual strategy P_{ind} and the population sex ratio P. It is obvious that the value of P is determined by a combined influence of all individual strategies in the population. However, this effect is not present in the classical approach. If an individual strategy $\hat{P}_{ind}$ will spread in the population, then we can expect convergence of P to the value of $\hat{P}_{ind}$. The Dusing-Fisher model analyzes only the effect of the population sex ratio P on the fitness of individual strategies; the effect of individual strategies on P is not considered. The second problem with the classical model is that it takes into account only female reproductive success. An individual strategy is interpreted as a decision process performed by a female. But since a fitness measure is assumed to be the total number of grandchildren, the gene transfer takes place through sons as well. This is an inconsistency because the fitness function is constructed for females only. It is possible that a male carrier of genes for a strategy P_{ind} will have a greater reproductive success than a female having this strategy. The question is: what influence the unconsidered effects described above will have on the predictions of the Dusing-Fisher model?

2. Multipopulation model. In order to address the above problems we need to construct an alternative model which would involve both males and females effects on the gene proliferation. The new model is very complicated from the mathematical point of view and therefore in this paper we limit ourselves to an overview of the new approach. It is based on a multi-population evolutionary game with a replicator dynamics (Argasiński [1]). We assume here that there are 3 different individual strategies in the population, for example $P_1 = 0.5$, $P_2 = 0.2$, and $P_3 = 0.8$. As in the classical approach, the individual strategy is the sex ratio (proportion of males) in a female brood. Now we construct a two-population model. The first subpopulation consists of males and the second of fe-

males. Individuals are distinguished by their sex and the strategy coded in their genes. A strategy is expressed only in females; male individuals are only its carriers. Current population state is described by

- $F = [f_1, f_2, f_3]$ – a vector of proportions of individual strategies in the female subpopulation,
- $M = [m_1, m_2, m_3]$ – a vector of proportions of individual strategies in the male subpopulation,
- P – a proportion of males in the population (or the population sex ratio).

The space of states of the population can be described as the product of two three-dimensional simplexes and a unit interval (one-dimensional simplex). Our model focuses on the dynamics of changes of the population state described by the replicator dynamics. Therefore, the fitness function does not describe the number of grandchildren but the number of direct descendants. The derivation of the fitness function is based on the computation of the probability of the number of descendants of an individual of a given sex and strategy. We assume that the production of a single descendant is a single Bernoulli trial, where the success means the transfer of the strategy gene to the new individual. For a male with the strategy P_i, the probability of such transfer to a male descendant is equal to $\theta_m^m(P_i) = 0.5\,(0.5f_1 + 0.2f_2 + 0.8f_3)$, an analogous value for a female descendant is $\theta_m^f(P_i) = 0.5\,(1 - (0.5f_1 + 0.2f_2 + 0.8f_3))$ (a haploid probability of a gene transfer multiplied by a probability of a proper sex choice according to the mean strategy of a potential mothers population). Analogously for a female we have $\theta_f^f(P_i) = 0.5\,(1 - P_i)$ and $\theta_f^m(P_i) = 0.5\,(P_i)$. A female makes x trials (a number of her kids) and a male individual makes $a_f x$ trials (a number of his kids), where a_f is the number of females per one male (analogously a_m is the number of males per one female, later on an additional upper index P_i will mean that we take into account only hosts of this strategy). Then we can calculate mean values of numbers of male and female descendants produced by a single male or a female parent with the strategy P_i denoted by $\mu_x^y(P_i)$, where the lower index describes the sex of a parent and the upper index describes the sex of an offspring. Hence for the strategy P_i we have $\mu_f^f(P_i) + a_m^{P_i}\mu_m^f(P_i)$ new female individuals per a single female parent, and an analogous value for males equals to $\mu_m^m(P_i) + a_f^{P_i}\mu_f^m(P_i)$. The mortality of parents may be treated as a negligible background fitness, so these values can be interpreted as per capita growth rates. According to Argasiński [1], these functions can be described in variables F, M and P. Therefore, depending on the sex of an individual, we obtain two kinds of the fitness (payoff) functions,

$$\begin{aligned} F_f(P_i, F, M, P) &= \frac{x}{2}\left((1 - P_i) + \frac{m_i}{f_i}\,(1 - (0.5f_1 + 0.2f_2 + 0.8f_3))\right) \text{ for females,} \\ F_m(P_i, F, M, P) &= x\frac{1 - P}{2P}\left(0.5f_1 + 0.2f_2 + 0.8f_3 + \frac{f_i}{m_i}P_i\right) \text{ for males.} \end{aligned} \tag{2}$$

These functions describe per capita growth rates of an individual characterized by the sex and an inherited strategy. Arguments of these functions are: the strategy P_i inherited by the individual and the population state described by vectors F, M and the value of P. Payoffs are interpreted as the per capita growth rate of an individual described by the

sex and its strategy at the moment when the population is in the state F, M, P. We can now derive the mean fitness function for the female subpopulation

$$\bar{F}_f(F,M,P) = f_1 F_f(P_1,F,M,P) + f_2 F_f(P_2,F,M,P) + f_3 F_f(P_3,F,M,P),$$

and the mean fitness function for the male subpopulation

$$\bar{F}_m(F,M,P) = m_1 F_m(P_1,F,M,P) + m_2 F_m(P_2,F,M,P) + m_3 F_m(P_3,F,M,P).$$

Subsequently, we can derive the mean fitness function for the whole population

$$\bar{F}(F,M,P) = P\bar{F}_m(F,M,P) + (1-P)\bar{F}_f(F,M,P).$$

We may now form the multi-population replicator equations (see [1]) which describe changes in proportions of strategies among females:

$$\dot{f}_i = f_i(F_f(P_i,F,M,P) - \bar{F}_f(F,M,P)); \quad i = 1,2,$$

changes in proportions of strategies among males:

$$\dot{m}_i = m_i(F_m(P_i,F,M,P) - \bar{F}_m(F,M,P)); \quad i = 1,2,$$

and the evolution of sex ratio:

$$\dot{P} = P(\bar{F}_m(F,M,P) - \bar{F}(F,M,P))$$

Then we obtain the following system of equations:

$$\begin{aligned}
&\dot{f}_i = x\left(\frac{f_i}{2}(1-P_i) + \left(\frac{m_i}{2} - f_i\right)(1-(0.5f_1+0.2f_2+0.8f_3))\right) \quad \text{for } i=1,2,\\
&\dot{m}_i = x\frac{1}{2}\left(\frac{1}{P}-1\right)(f_iP_i - m_i(0.5f_1+0.2f_2+0.8f_3)) \quad \text{for } i=1,2,\\
&\dot{P} = x(1-P)(0.5f_1+0.2f_2+0.8f_3-P).
\end{aligned} \tag{3}$$

Solutions of the above equations show that:

• Natural selection does not favor the individual strategy 0.5 (or any other strategy). Trajectories of the system converge to restpoints. For every rest point, the mean strategy of the female subpopulation (that is the primary sex ratio) is equal 0.5 and the proportion of carriers of a strategy P_i must satisfy the following condition: $f_iP_i = m_i(0.5f_1+0.2f_2+0.8f_3)$.

• A trajectory of the population sex ratio (the secondary sex ratio) P can pass the value 0.5, which is stable from the point of view of the classical theory. P is attracted by the current value of the primary sex ratio and follows it until the whole system reaches a stable state. This happens when the primary sex ratio (and in effect the secondary sex ratio) is equal to 0.5.

• Changes of initial conditions of the male subpopulation may affect the behavior of the whole system.

3. Conclusions. Our model shows that superiority based on the number of grandchildren of a female cannot be extrapolated to future generations. Trajectories of the evolution of the male subpopulation may strongly affect the behavior of the whole system. Our major conclusion is that the role of male hosts of sex-ratio genes has been completely disregarded in the classical theory. Numerical solutions of replicator equations show that

dynamics of this system is very complex. Continuation of this work may provide new interesting results.

References

[1] K. Argasiński, *Dynamic multipopulation and density dependent evolutionary games related to replicator dynamics. A metasimplex concept*, Mathematical Biosciences 202 (2006), 88–114.

[2] I. M. Bomze and B. M. Potscher, *Game Theoretical Foundations of Evolutionary Stability*, Lecture Notes in Economic and Mathematical Systems 324, Springer, 1989.

[3] R. Cressman, *The Stability Concept of Evolutionary Game Theory*, Springer, 1992.

[4] R. A. Fisher, *The Genetical Theory of Natural Selection*, Oxford University Press, 2000.

[5] D. J. Futuyma, *Evolution*, Sinauer Associates, 2005.

[6] *Encyclopedia of Ecology and Environmental Management*, P. Calow (ed.), Blackwell, 1999.

[7] J. Hofbauer and K. Sigmund, *The Theory of Evolution and Dynamical Systems*, Cambridge University Press, 1988.

[8] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics*, Cambridge University Press, 1998.

[9] D. C. Queller, *Sex Ratios and Social Evolution*, Current Biology 16 no. 17.

[10] M. Ridley, *Evolution*, Blackwell, 1996.

[11] J. Weibull, *Evolutionary Game Theory*, MIT, 1995.

STOCHASTIC MODELS IN BIOLOGICAL SCIENCES
BANACH CENTER PUBLICATIONS, VOLUME 80
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WARSZAWA 2008

RELATIONSHIP BETWEEN THE SELECTION PRESSURE AND THE RATE OF MUTATION ACCUMULATION

KATARZYNA BOŃKOWSKA, PRZEMYSŁAW BIECEK,
AGNIESZKA ŁASZKIEWICZ and STANISŁAW CEBRAT

Department of Genomics, Wrocław University
Przybyszewskiego 63/77, 51-148 Wrocław, Poland
E-mail: cebrat@smorfland.uni.wroc.pl

Abstract. We use the diploid, sexual Penna ageing model and its modification with noise and environment fluctuations to analyse the influence of random death on the accumulation of defective genes in the genetic pool of populations evolving under different environmental conditions.

1. Introduction. Medawar [4] claimed that the force of natural selection weakens with increasing age. A weaker selection pressure enables the accumulation of mutations in the individuals' genomes in the late expressed genes. Consequences of genetic defects, which appear late in the lifetime, may be unimportant from the evolutionary point of view since they affect a small number of individuals. Nevertheless, they shorten the individuals' lifespans. This phenomenon is known as the mutation accumulation theory of ageing.

The Penna model [5] confirms this theory. The model assumes the chronological activation of genes during individuals' lives, which leads to the accumulation of defective genes expressed late during life. For all organisms, living in the natural world is connected with inevitable deaths due to restrictions of resources or diseases and predators. In the Penna model, the size of the population is regulated by a parameter which can operate only on newborns or the whole population. In the first strategy, there are no random deaths during lifetime—the Medawar's assumption, that random deaths are responsible for the weaker selection of individuals due to their older age, disappears. The gradient of accumulated defects and specific age structure of population still persists [3].

2000 *Mathematics Subject Classification*: Primary 92D25; Secondary 92D15.

Key words and phrases: Penna model, ageing, mutation, selection, population evolution.

This work was supported by the Poland's State Committee for Scientific Research, grant # 105/E-344/SPB and Polish Foundation for Science. It was done in the framework of European programs: COST Action P10 and NEST - GIACS.

The paper is in final form and no version of it will be published elsewhere.

 © Instytut Matematyczny PAN, 2008

In the modified Penna model [1] the health status of an individual is described by variance of its fluctuations. Each expressed deleterious mutation increases the fluctuations. Additionally, the fluctuations of the environment are superimposed on the fluctuations of individuals in the population. The individual dies if the combination of both stochastic processes trespasses the limit (level of homeostasis) set as the model parameter. The genes are switched on chronologically leading to the accumulation of defective genes expressed during the late periods of life in the genetic pool of the population. A decrease of the variation of the environmental noise, corresponding to the decrease in the selection pressure, increases the average expected lifespan of individuals.

2. Model. In the sexual version of the Penna model, individuals are represented by two strings of bits of declared length L. The genes are represented by single bits. There are two possible states for a gene. 0 means that the gene is correct, while a mutated one is represented by 1. We have assumed that all defects are recessive. There are no reversions.

In the model genes are switched on chronologically which means that in the first Monte Carlo (MC) step after birth, the first locus of the genome is expressed and checked whether it is correct or not. In the second MC step, the second locus is expressed and so on. If the number of expressed defective loci reaches the value of a parameter T - the individual dies. Additionally, each individual may be killed by the environmental restrictions. In the model, the space and food limits are introduced as the Verhulst factor:

$$V = 1 - N_t/N_{max}, \tag{1}$$

where V is a probability of survival, N_t is a current population size while N_{max} is a maximum population size, called also the maximum capacity of the environment.

If the individual stays alive until the minimum reproduction age R, it generates an offspring with declared probability B. In the female genome, one locus is randomly chosen with probability C, bitstrings at this locus are cut and homologous fragments are exchanged. A new mutation is introduced to the gamete with probability M. Then, the male individual is randomly chosen and the same processes happen during its gamete production. The zygote is formed by joining female and male gametes. Sex of the newborn is established with an equal probability for male and female.

Let N_{Rt} denote the number of individuals in the current population in the reproduction age while $N_{ft} = \max(0, N_{max} - N_t)$ denotes the number of "free slots" in the population. The number of newborns follows the distribution

$$\mathcal{B}\left(\mathcal{B}(N_{Rt}, B), \frac{N_{ft}}{N_{max}}\right) \underset{N_{max}\to\infty}{\longrightarrow} \mathcal{N}\left(N_{Rt} * \frac{B * N_{ft}}{N_{max}}, N_{Rt} * \frac{B * N_{ft}}{N_{max}} * \left(1 - \frac{B * N_{ft}}{N_{max}}\right)\right), \tag{2}$$

where $\mathcal{B}$ signifies the binomial and $\mathcal{N}$ the Gaussian distribution.

Thus, the expected number of newborns is $N_{Rt} * \frac{B * N_{ft}}{N_{max}}$. If $N_{Rt} << N_{max}$ the population will become smaller and smaller and in the end it may extinct.

Because the Penna model assumes chronological activation of genes during an individual's lifespan, it leads to the accumulation of defective genes expressed late during the lifespan. Genes switched on before reproductive age are under strong selection pressure and the fraction of defects in this set of genes is low. Genes switched on after the mini-

mum reproductive age are under weaker selection pressure and an increasing fraction of defects is observed.

To introduce the role of the environment, we modified the diploid Penna model in which each individual is represented by a pair of bitstrings. The state of the organism and the environment are determined by the value of their fluctuation [1];

$$P_i(t) \sim \mathcal{N}(0, \sigma_i^2(t)), \qquad E(t) \sim \mathcal{N}(0, \sigma_e^2). \tag{3}$$

The amplitude of the fluctuation depends on the genetic background of the organism

$$\sigma_i^2(t) = \sigma_0^2 + \sum_{j=1}^{age(i,t)} g_{ij}\sigma_d^2. \tag{4}$$

where σ_d is the increase of variation caused by a defective gene; σ_0 is the base variation of the individual state (base penalty) while g_{ij} is the number of defective pairs of alleles in loci j for individual i. Thus, the more defective loci the higher is the variance of the individual's state and higher the risk of the individual's death, since individual i is killed if the sum of both states crosses the homeostasis limit F [1];

$$P_i(t) + E(t) \geq F. \tag{5}$$

There is a similarity between parameter T in the standard Penna model and parameter F in the considered model. Both parameters define the maximum acceptable harm in the individuals, in the standard Penna model it is simply the number of allowed defects, since in the modified model it is the maximum deviation from the expected environmental state.

3. Results and discussion. As it was mentioned above, the role of the Verhulst factor is to prevent the unlimited growth of the population. The Verhulst factor is responsible for the random deaths of individuals. In the first version, the Verhulst factor operates during the whole life independently of age (strategy $Vall$). Nevertheless, these deaths are not completely random considering the genetic structure of individuals. Individuals with a better genetic structure (better fitted to the environment) live longer and they are exposed to the Verhulst factor for a longer time. If the Verhulst factor determines the survival of newborns only (strategy Vb), the individuals are killed with the same probability, independently of their genetic structure. This strategy generates a better genetic structure of the population than the $Vall$ strategy [2] (Fig. 1). We have also observed longer life spans of individuals' and higher fraction of individuals in the reproductive age (Fig. 2).

In the noisy model, the state of the environment randomly fluctuates, thus, even individuals with very good phenotypes may die, but individuals with defective alleles die more frequently due to the higher variation of their states. Therefore, this randomness and the one represented by the Verhulst factor are different. The noisy Penna model allows us to investigate the role of the environment on the population evolution.

We considered two populations, one which exists in the mild environment and one living in the more severe environment. The difference between populations occurs through

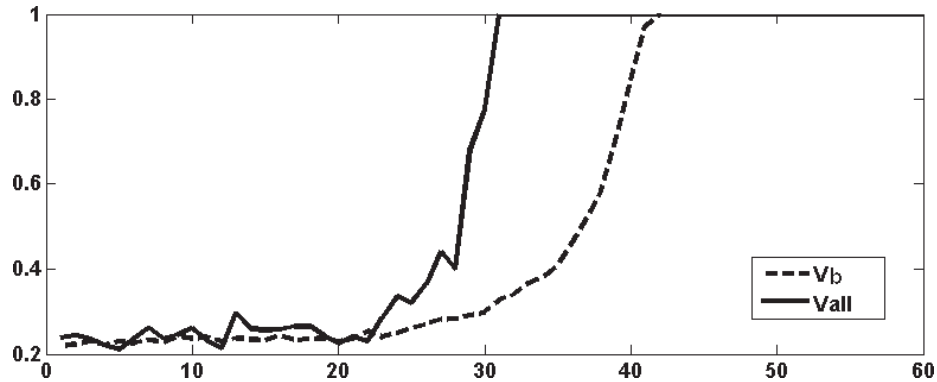


Fig. 1. Fraction of defective genes observed if the Verhulst factor operates only on newborns (Vb) or on the entire population ($Vall$). Parameters of simulations: $L = 64$, $R = 20$, $B = 1$, $T = 3$.

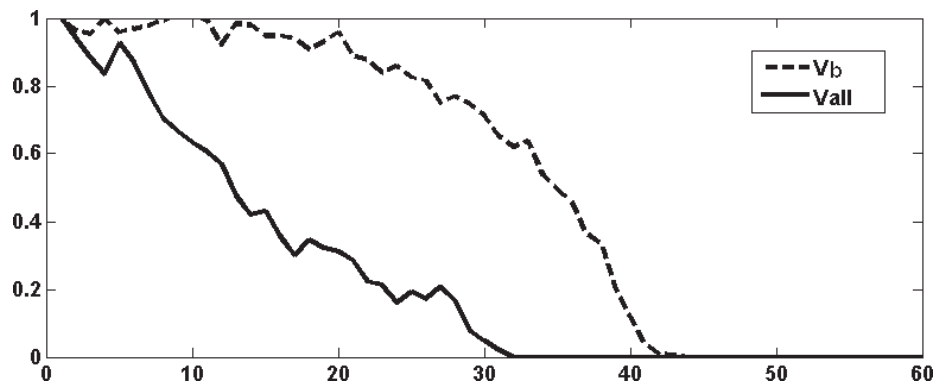


Fig. 2. Age distribution generated by Verhulst factor operating only on newborns (Vb) or on the entire population ($Vall$). Parameters of simulations are as in Fig. 1.

values of σ_e and σ_d. In the more severe environment both values are larger, thus the individuals with the same number of defects die more often in the unfriendly environment.

The distributions of defects in the genomes are presented in Fig. 3. The average lifespan for a population in the friendly environment is longer due to the lower probability of death. On the other hand, the population in the severe environment has a better genetic structure, i.e. fewer defective genes are expressed before reproduction age.

We checked the results of competition between the two populations. After 100000 MC steps, during which these populations evolve separately, we allowed individuals to migrate between the environments. With probability 0.01 each individual could decide to switch its environment and live in different environment than their parents. Mating between individuals from the mild and the severe environments was forbidden, thus, we were able to check which population was better adapted. We expected that the population from the mild environment had smaller chance to survive in the unfriendly conditions, but this population lost (i.e. all individuals in the population died after 100000 MC steps) even

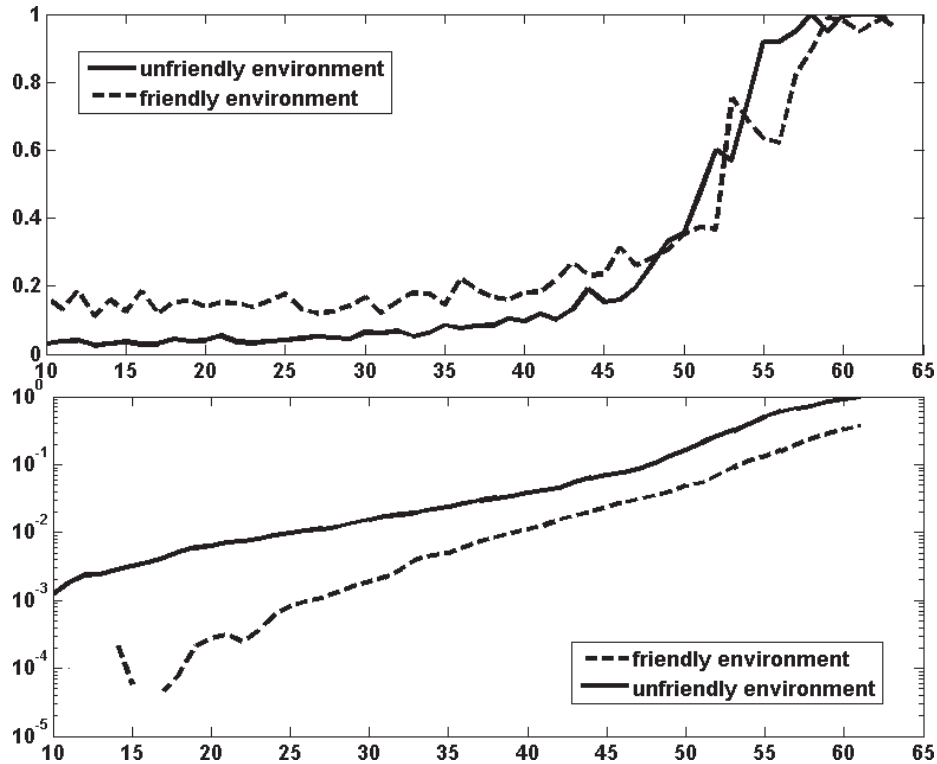


Fig. 3. The distribution of defective genes (top) and mortality (bottom) in populations living in friendly and unfriendly environments.

in the friendly environment in which it was used to live. We repeated these simulations 100 times and in 99 cases the population from friendly environment was eliminated from both environments.

4. Conclusions. Our model confirms the trivial rules that the structures of individuals' genomes in the evolving populations are the results of two counterbalancing forces: selection and mutation pressure. Nevertheless, environment can eliminate individuals randomly or selectively. Random eliminations decrease the speed of adaptation of population and its evolution and its negative impact depends on the age of affected individuals. The most efficient way for the evolution of the fittest genomes is sacrificing the youngest individuals in the random process and imposing the directional selective force on the older members of population.

References

[1] P. Biecek and S. Cebrat, *Fluctuations, environment, mutations accumulation and ageing*, Int. Journal of Modern Physics C, 6 (17) (2006), 923–931.

[2] A. Łaszkiewicz, *Biological interpretation of the Penna model parameters*, PhD dissertation, 2006.
[3] J. S. Martins and S. Cebrat, *Random deaths in a computational model for age-structured populations*, Theory in Bioscien. 119 (2000), 156–165.
[4] P. B. Medewar, *An unsolved problem of biology, an inaugural lecture delivered at University College London*, H. K. Lewis, London, 1952.
[5] T. J. P. Penna, *A bit-string model for biological aging.* J. Stat. Phys. 78 (1995), 1629–1633.

STOCHASTIC MODELS IN BIOLOGICAL SCIENCES
BANACH CENTER PUBLICATIONS, VOLUME 80
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WARSZAWA 2008

THERMODYNAMICS OF DNA MICROARRAYS

ENRICO CARLON

Interdisciplinary Research Institute of Lille
École Polytechnique Universitaire de Lille (Polytech'Lille)
Cité Scientifique, F-59652 Villeneuve d'Ascq, France
E-mail: enrico.carlon@polytech-lille.fr

Abstract. DNA microarrays have been widely used in molecular biology laboratories. The main current application of these devices is the determination of the gene expression level for thousands of genes simultaneously. Here we review a recently introduced physical model for hybridization (i.e. the binding of complementary DNA strands) in Affymetrix arrays and compare it to experimental results. The experimental data follow rather well the microscopic model and the approach offers several advantages compared to traditional methods of microarray data analysis.

1. Introduction. A gene is a small fragment of DNA which contains the code for the synthesis of a specific protein (for simplicity we limit ourselves here to discuss the protein coding genes, as there are indeed genes that do not encode for a protein). The process leading to the synthesis of a protein consists of two steps: the gene is first transcribed into messenger RNA (mRNA) which is then translated into protein.

The 30,000 human genes are not expressed, i.e. transcribed into mRNA, at all times. Some genes are tissue specific in the sense that they are expressed only in some given tissues, while others are more ubiquitous and are expressed in all tissues. For instance: pancreatic cells specialize in the synthesis of proteins as trypsin, which is an enzyme involved in the digestion and a main component of the so-called pancreatic juice. The associated gene is highly expressed in pancreatic cells, while is typically not expressed in cells of other tissues. This information alone suggests that trypsin is a protein which is involved in some basic pancreatic function. Knowing when, in which amount and in which tissue a given gene is expressed suggests the functions of the corresponding protein. This is a very valuable information to understand the function of our genes: currently the function of only about 20% of the 30,000 human genes has been identified.

2000 *Mathematics Subject Classification*: Primary 92E20; Secondary 80A10.
Key words and phrases: DNA microarrays, DNA hybridization.
The paper is in final form and no version of it will be published elsewhere.

 © Instytut Matematyczny PAN, 2008

DNA microarrays are devices which, in principle, can measure the gene expression level of thousands of genes simultaneously. They are made of a small glass plate containing single stranded DNA fragments anchored at the surface. A microarray experiment consists of the following steps: first a class of cells is extracted from a tissue and the total mRNA from those cells is isolated and marked with fluorescent molecules. The solution obtained is then put in contact with the microarray. If a given sequence floating in solution (known as a *target*) meets its complement sequence anchored at the glass surface (the *probe*) it binds. The binding of complementary probe-target sequences is called hybridization. As the target strand is labeled with fluorescent molecules the fraction of hybridized probes can be measured by optical methods. Fig. 1(left) shows an image of an Affymetrix (a leading company in the DNA microarrays market) array.

Here we will review some recent results of an analysis of DNA microarray data using physical-chemistry models. We will focus on microarrays produced by Affymetrix. These arrays have two special features compared to all other arrays: 1) Multiple probes per gene are used 2) for each probe perfectly complementary to DNA in solution there is an additional probe with a sequence differing by a single nucleotide. Thus, one distinguishes between perfect matching (PM) probes and mismatching (MM) ones.

In the next sections we describe the model and the analysis of the experimental data. Several other models have been proposed in the recent literature (for a review see [Halperin] and references therein). The main features of the model presented here are 1) the use of the hybridization free energies for RNA/DNA duplexes, 2) the inclusion of the mismatches in the analysis, 3) the use of very few free fitting parameters and 4) the inclusion of the hybridization between partially complementary strands in solution.

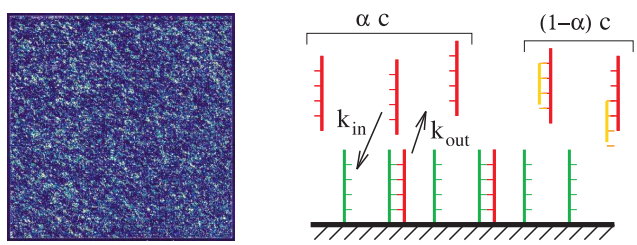


Fig. 1. Left: Fluorescent image of a scanned Affymetrix arrays. The image is of an array containing a grid of $640 \times 640 = 409,600$ different probes on a surface of 1.28×1.28 cm^2. In the latest Affymetrix chips (2006) the same surface area can contain up to 6.4 millions different sequences. Right: Hybridization model developed in [Carlon]. 1) probes with a specific DNA sequence hybridize with complementary molecules in solution with attachment/detachment rates $k_{\rm in}$ and $k_{\rm out}$. 2) Target molecules can partially hybridize in solution leading to an effective reduction of the target concentration to a factor αc.

2. The extended Langmuir model. As the Langmuir model in its basic form does not seem to fit sufficiently well the experimental microarrays data (see [Carlon]) we have modified it to include the possibility of hybridization between targets in solution.

According to this *extended* model the fluorescent intensity measured for a given probe is [Carlon]

$$I = I_0 + \frac{A\alpha c e^{\beta\Delta G}}{1 + \alpha c e^{\beta\Delta G}} \tag{1}$$

where c is the target concentration, ΔG is the hybridization free energy, $\beta = 1/RT$ the inverse temperature (R the gas constant), A a scale factor and I_0 the non-specific background signal. The concentration c is a measure of the gene expression level that one wants to determine. The difference with the basic Langmuir model is the inclusion of a factor α, which takes into account of the effective reduction of the target molecules in solution [Carlon], as illustrated in Fig. 1. In the model we handle PM and MM signals on the same footing: perfect matches hybridize more efficiently to the probes than mismatches and this difference results in a different hybridization free energy ΔG. The latter is calculated from experimental data for RNA/DNA hybrids taken from [Sugimoto] using the nearest neighbor model. The hybridization in solution is taken into account in a kind of mean-field approximation and the sequence-dependent factor α contains two adjustable parameters [Carlon]. Two other fitting parameters in Eq. (1) are A and β: the parameter A sets the overall level of fluorescence, given in arbitrary units by Affymetrix and β is an inverse effective temperature. The latter is higher than the experimental temperature suggesting that the interaction strength in hybridization with surface-bound DNA, as in Affymetrix microarrays, is overall lower than the corresponding interaction between two strands hybridizing in solution. For a detailed discussion of this effect see [Carlon].

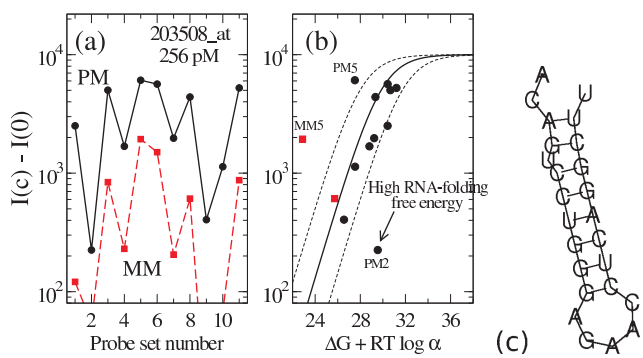


Fig. 2. (a) Plot of the background-subtracted intensity of the Affymetrix spike-in experiment for perfect matches (PM) and mismatches (MM) probes of the probe set 203508_at (corresponding to the gene 1B of the tumor necrosis factor superfamily), which in the experiment is spiked at 256 pM. (b) The same data plotted as function of the hybridization free energy according to Eq. (1) (similar plots can be produced with a free software SpikeLI, available under the Bioconductor project [SpikeLI]). The solid line is the predicted Langmuir isotherm with $c = 256$ pM. The dotted lines correspond to concentrations four times higher and smaller. (c) Minimal free energy configuration for the target sequence in solution which is complementary to probe 2. The low signal PM2 observed in (b) is therefore very likely caused by the folding of the sequence in solution.

3. Data analysis. We have analyzed freely available Affymetrix data of two Latin square experiments. In these experiments some targets are spiked-in at a known concentration ranging from 0.125 up to 1024 pM (picomolar), covering all the concentrations of biological interest. As c is known in Eq.(1) we have used the data to fit the other four global fitting parameters. Fig. 2 gives an example of the signal measured from a probe set, ie the collection of all probes PM and MM corresponding to a given gene. In the example the probe set is labeled 203508_at (Affymetrix nomenclature) which corresponds to the gene 1B of the tumor necrosis factor receptor superfamily. In Fig. 2(a) we plot the PM and MM intensities versus the probe number. In the specific experiment shown the concentration is 256 pM. Note the PM signals are higher than the MM signals, consistent with thermodynamics expectations. There is a remarkable variability within each set: for instance the PM signals vary from 200 to 2000 (Affymetrix fluorescence units). Fig. 2(b) shows the same data points as in (a) plotted as function of the variable $\Delta G + RT \log \alpha$. The solid line is the Langmuir isotherm corresponding to a concentration of 256 pM. As seen in the figure, the experimental data align quite well along the theoretical prediction. The dotted lines correspond to two Langmuir isotherms at concentrations 64 and 1024 pM, i.e. four times bigger and smaller than the actual experimental concentration. The graph shows the presence of two "outliers" corresponding to data out of the expected trend: these are probes 2 and 5. The target sequence 2 turns out to be highly self-complementary (Fig. 2(c)): a calculation of folding free energies [Heim] shows that the probability of forming a secondary structure is high, which explains probably why the probe 2 has a signal much lower than the predicted Langmuir curve.

Another way of analyzing the experimental data is through a scaling collapse plot. The form given in Eq. (1) suggests that when using a scaling variable $x' = \alpha c \exp(\beta \Delta G)$ all data should collapse onto a single universal master curve of the form $Ax'/(1 + x')$. This is indeed seen in the plots of Fig. 3 which shows some collapse plots for four different probe sets: each plot contains about 300 experimental data points with concentrations ranging from 0.125 to 512 pM. There is a global trend to alignement onto a single curve. Note that only few probes deviate from the theoretical curves as for instance probes 9 and 4 of the probe set 205398_s_at and probe 2 of the probe set AFFX-r2-TagA_at. In Fig. 3 the probe sets 205398_s_at 207655_s_at contain sequences complementary to human genes, while the probe sets AFFX-r2-TagE_at and AFFX-r2-TagA_at contain sequences of artificial origins. The latter spikes were synthesized by Affymetrix and contain sequences which are believed to be unique within the human genome. Our analysis [Heim] shows that the quality of the collapses is the best for sequences of artificial origin. This suggests that the deviations from the Langmuir model have probably biological, rather than thermodynamic origin, i.e. are due to the non-optimal selection of probes.

4. Conclusions. In this paper we reviewed some recent results about the analysis of Affymetrix microarrays data using a physical-chemistry based model [Carlon]. When appropriately rescaled the experimental data tend to collapse to a single universal curve, in agreement with Langmuir thermodynamics. This approach provides an alternative way

for determining the gene expression level (i.e. the target concentration c) compared to traditional bioinformatic tools currently available.

The present results were obtained in collaboration with Thomas Heim, Léon-Charles Tranchevent and Gerard Barkema. We acknowledge financial support from the Van Gogh Programme d'Actions Intégrées (PAI) 08505PB of the French Ministry of Foreign Affairs.

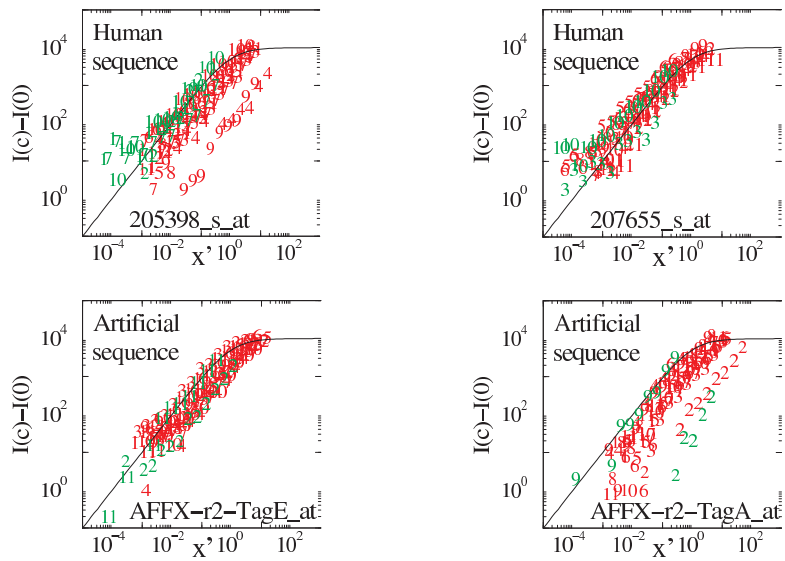


Fig. 3. Collapse plots for 4 different probe sets of the Affymetrix "Latin square" spike-in experiments. The solid line is the Langmuir isotherm (Eq. (1)) plotted as a function of the scaling variable $x' = \alpha c \exp(\beta \Delta G)$.

References

[Carlon] E. Carlon and T. Heim, Physica A 362 (2006), 433.

[Halperin] A. Halperin A. Buhot and E.B. Zhulina, J. Phys.: Cond. Matt. 18 (2006), S463.

[Heim] T. Heim, L.C. Tranchevent, E. Carlon and G.T. Barkema, J. Phys. Chem. B 110 (2006), 22786.

[SpikeLI] D. Baillon, P. Leclercq, S. Ternisien, T. Heim and E. Carlon, SpikeLI: A Bioconductor package for the analysis of Affymetrix spike-in data, http://www.bioconductor.org/packages/1.9/bioc/html/spikeLI.html

[Sugimoto] N. Sugimoto *et al.*, Biochemistry 34 (1995), 11211.

[Xia] T. Xia *et al.*, Biochemistry 37 (1998), 14719.

STOCHASTIC MODELS IN BIOLOGICAL SCIENCES
BANACH CENTER PUBLICATIONS, VOLUME 80
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WARSZAWA 2008

ADAPTIVE DYNAMICS IN LOGISTIC BRANCHING POPULATIONS

NICOLAS CHAMPAGNAT
TOSCA team
INRIA Sophia Antipolis–Méditerrannée
2004 Route des Lucioles, BP 93, 06902 Sophia Antipolis, France
E-mail: Nicolas.Champagnat@sophia.inria.fr

AMAURY LAMBERT
Unit of Mathematical Evolutionary Biology
CNRS UMR 7625 Laboratoire d'Écologie et Évolution
École Normale Supérieure
46 rue d'Ulm, 75230 Paris Cedex 05, France
E-mail: amaury.lambert@ens.fr
and
CNRS UMR 7625 Laboratoire d'Écologie et Évolution
UPMC Univ. Paris 06
7 quai Saint Bernard, 75252 Paris Cedex 05, France

Abstract. The biological theory of adaptive dynamics proposes a description of the long-time evolution of an asexual population, based on the assumptions of large population, rare mutations and small mutation steps. Under these assumptions, the evolution of a quantitative dominant trait in an isolated population is described by a deterministic differential equation called 'canonical equation of adaptive dynamics'. In this work, in order to include the effect of genetic drift in this model, we consider instead finite, randomly fluctuating populations and weak selection.

We consider a trait-structured population subject to mutation, birth and competition of logistic type, where the number of coexisting types may fluctuate. Applying a limit of rare mutations to this population while keeping the population size finite leads to a jump process, the so-called 'trait substitution sequence', where evolution proceeds by successive invasions and fixa-

2000 *Mathematics Subject Classification*: Primary 92D15; Secondary 60K35, 60J85, 60J70, 92D10, 60J75, 92D25, 92D40, 60J80, 60J25.

Key words and phrases: logistic branching process, multitype birth death competition process, density-dependence, fixation probability, genetic drift, weak selection, adaptive dynamics, invasion fitness, timescale separation, trait substitution sequence, diffusion approximation.

The paper is in final form and no version of it will be published elsewhere.

 © Instytut Matematyczny PAN, 2008

tions of mutant types. The probability of fixation of a mutant is interpreted as a fitness landscape that depends on the current state of the population. Rescaling mutation steps (weak selection) then yields a diffusion on the trait space christened 'canonical diffusion of adaptive dynamics', in which genetic drift (diffusive term) is combined with directional selection (deterministic term) driven by the fitness gradient.

Finally, in order to compute the coefficients of this diffusion, we seek explicit first-order formulae for the probability of fixation of a nearly neutral mutant appearing in a resident population. The first-order term is a linear combination of products of functions of the initial mutant frequency times 'invasibility coefficients' associated with fertility, defence, aggressiveness and isolation, which measure the robustness (stability with respect to selective strengths) of the resident type. Some numerical results on the canonical diffusion are also given.

1. Introduction. The recent biological theory of adaptive dynamics [6,9] proposes a description of the long term evolution of an asexual population by putting emphasis on the ecological interactions between individuals, in contrast with classical population genetics models which focus on the genetic structure of the population. The basic models are individual-based models in which the population dynamics is precisely described and includes birth, death, competition and mutation. The basic idea of the theory of adaptive dynamics is to try to get insights about the interplay between ecology and evolution by studying the invasion of a single mutant type appearing in a simplified (monotype stable) resident population. The evolution of the population can then be described as a sequence of mutant invasions in the population. If the resident type goes extinct when the mutant type invades (we say that the mutant type *becomes fixed*), the evolution is described by the so-called 'trait substitution sequence' (TSS) [10]. This approach has revealed powerful to predict the qualitative behaviour of complex evolutionary dynamics. In particular, it allows one to determine the (local) direction of evolution in the space of phenotypic traits (or simply *traits*) from the individual ecological parameters, and to predict and explain the phenomenon of evolutionary branching [10], where a population, initially composed of individuals with traits concentrated around a single trait value, divides into two (or more) subpopulations concentrated around distinct trait values that stably coexist because of their ecological interactions. The description of this phenomenon is an important achievement of this theory, as well as the 'canonical equation of adaptive dynamics' [3], which describes the evolution of the dominant trait of the population as a deterministic "hill-climbing" process on a fitness landscape which depends on the current state of the population (see (1) below).

More formally, as soon as eternal coexistence of two or more types is not permitted by the model, the evolution can be described by the sequences $(T_n)_n$ and $(V_n)_n$, where T_n is the n-th time where the population becomes monomorphic (i.e. composed of only one type) and V_n is the surviving type at time T_n. The sequence (V_n) is the above-mentionned TSS. It is possible to prove the convergence of an individual-based model to the TSS under two biologically motivated assumptions [10,1]. First, the *assumption of rare mutations* guarantees that, in the timescale of mutations, the widths of time intervals during which the population is polymorphic vanish, so that there is only one type surviving at any time t. To prevent the population from rapidly becoming

extinct in the new timescale, one also has to rescale population sizes, thereby making the *assumption of large populations*.

Subsequently, the TSS is a Markov jump process on the trait space whose semigroup is shown [1] to depend on the *invasion fitnesses* ([9]) $f(x,y)$, $x,y \in \mathcal{X}$, where $f(x,y)$ is defined as the *expected growth rate* of a single individual of type y — the mutant — entering a monomorphic population of type x 'at equilibrium' — the residents. Note that this fitness is not given *a priori*, but derived from the microscopic model of individual interactions. Because of the assumption of large population, the sign of this fitness determines the possibility of invasion of a mutant type: if $f(x,y) < 0$, the mutant type y cannot invade a resident population of type x. Thus, evolution proceeds by *successive invasions of (only) advantageous mutant types* replacing the resident one.

The 'canonical equation of adaptive dynamics' [3], which describes the evolution of a one-dimensional trait x as the solution of the following ODE, is obtained from the TSS in the limit of *small mutations*:

$$\frac{dx}{dt} = \frac{1}{2}\sigma(x)^2\mu(x)\bar{n}(x)\frac{\partial}{\partial y}f(x,x), \tag{1}$$

where $\sigma(x)^2$ stands for the (rescaled) variance of the mutation step law, $\bar{n}(x)$ for the equilibrium size of a pure x-type population, and $f(x,y)$ for the invasion fitness mentioned above. Note how only advantageous types get fixed (the trait follows the fitness gradient) and how the fitness landscape $y \mapsto f(x,y)$ depends on the current state x of the population.

However, it is well-known that slightly deleterious types can be fixed by chance in finite populations. This phenomenon is known under the name of *genetic drift*. Depending on the strength of genetic drift, selection is said to be *strong* (genetic drift has negligible effects) or *weak*. In the large population asymptotic from which the TSS of adaptive dynamics is derived, genetic drift has negligible impact compared to the action of selection. Therefore, the fixation of slightly deleterious types cannot be observed. Our goal here is to include genetic drift in the adaptive dynamics models by considering *finite populations* under *weak selection*. We continue using the bottom-up approach of adaptive dynamics; that is, model (macroscopic) evolution from (microscopic) populations. In particular, we allow the population sizes to fluctuate randomly through time and we aim to reconstruct a fitness function from the microscopic parameters.

After the description of the model (Section 2), we derive a new TSS in the limit of rare mutations (Section 3), from which a limit of small mutations gives what we call the 'canonical diffusion of adaptive dynamics' (Section 4). The coefficients of this diffusion involve the first-order derivatives of the fixation probabilities, which are computed in Section 5 as a linear combination of four fundamental components associated to *fertility*, *defence*, *aggressiveness* and *isolation*. New numerical results on the robustness of the population with respect to these fundamental components are also given, as well as some consequences on the canonical diffusion of adaptive dynamics in large populations.

2. The microscopic model. We will restrict here to logistic interaction. More general models are considered in [2].

A monotype (binary) logistic branching process (LBP, see [7]) with dynamical parameters (b,c) is a Markov chain in continuous time $(X_t; t \geq 0)$ with nonnegative integer values and transition rates

$$q_{ij} = \begin{cases} bi & \text{if } j = i+1, \\ ci(i-1) & \text{if } j = i-1, \\ -i(b+c(i-1)) & \text{if } j = i, \\ 0 & \text{otherwise.} \end{cases}$$

The non-linear term $ci(i-1)$ describes competition mortality due to random encounters between individuals. Other terms correspond to independent birth events with constant individual rates. This Markov chain is positive-recurrent and converges in distribution to a r.v. ξ, where ξ is a Poisson variable of parameter $\theta := b/c$ conditioned on being nonzero

$$\mathbb{P}(\xi = i) = \frac{e^{-\theta}}{1-e^{-\theta}}, \frac{\theta^i}{i!}, \qquad i \geq 1. \tag{2}$$

We consider a multitype asexual birth and death process with mutation, generalizing this LBP. At any time t, the population is composed of a finite number $N(t)$ of individuals characterized by their phenotypic traits $x_1(t), \ldots, x_{N(t)}(t)$ belonging to a given trait space $\mathcal{X}$, assumed to be a closed subset of $\mathbb{R}^k$. The population state at time t is represented by the counting measure on $\mathcal{X}$

$$\nu_t = \sum_{i=1}^{N(t)} \delta_{x_i(t)}.$$

The population dynamics is governed by the following parameters.

- $b(x)$ is the rate of birth from an individual of type x. The function b is assumed to be $\mathcal{C}_b^2$.
- $c(x,y)$ is the rate of death of an individual of type x due to the competition with another individual of type y. Therefore, the total death rate of an individual of type x in a population ν may be written as $\int c(x,y)(\nu(dy) - \delta_x(dy))$. In this expression, the Dirac mass at x substracted to the measure ν means that the individual does not compete with himself. The function c is assumed to be $\mathcal{C}_b^2$ and bounded away from 0 on $\mathcal{X}^2$.
- $\gamma\mu(x)$ is the probability that a birth from an individual with trait x produces a mutant individual, where $\mu(x) \in [0,1]$ and where $\gamma \in (0,1]$ is a parameter scaling the frequence of mutations. When there is no mutation, the new individual inherits the trait of its progenitor. In Section 3, we will be interested in the limit of rare mutations ($\gamma \to 0$).
- $M(x,dh)$ is the law of the trait difference $h = y - x$ between a mutant individual with trait y and the trait x of his progenitor. We assume that $M(x,dh)$ has 0 expectation (no mutation bias), i.e. $\int hM(x,dh) = 0$, and has a density on $\mathbb{R}^k$ which is uniformly bounded in $x \in \mathcal{X}$ by some function $\bar{M}(h)$ with finite third-order moment.

We will denote the dependence of ν_t on the parameter γ with the notation ν_t^γ. Observe that such a population cannot become extinct because the death rate is 0 when there is only one individual in the population. Since we want to apply a limit of rare mutations

while keeping the population size finite, this is necessary to prevent the population to become extinct before any mutations occur.

Let $\xi(x)$ be a random variable whose law is the stationary distribution of a pure x-type population with **no** mutation ($\mu \equiv 0$). This law is given by (2) where θ is replaced by $\theta(x) := b(x)/c(x,x)$.

The last notation needed concerns a population with initially only two types x and y and with **no** mutation. Then $\nu_t = X_t\delta_x + Y_t\delta_y$, where $(X_t, Y_t : t \geq 0)$ is a bivariate Markov chain. For this Markov chain, $\mathbb{P}(T < \infty) = 1$, where T is the first time where either X_t or Y_t reach 0. We call *fixation* (of the mutant y) the event $\{X_T = 0\}$. The probability of fixation will be denoted by $u_{n,m}(x,y)$

$$u_{n,m}(x,y) := \mathbb{P}(X_T = 0 \mid X_0 = n, Y_0 = m).$$

3. The trait substitution sequence in finite populations. In this section, we apply the limit of rare mutations ($\gamma \to 0$) to the process ν^γ, in order to describe the evolution of the population as a TSS in finite population. This limit requires to rescale time properly, as t/γ, to describe the evolution on the mutation timescale.

THEOREM 1. *Fix $x \in \mathcal{X}$. Assume that $\nu_0^\gamma = N_0^\gamma \delta_x$ where $\sup_{\gamma\in(0,1)} \mathbb{E}((N_0^\gamma)^p) < \infty$ for some $p > 1$. Then, for any $0 < t_1 < \ldots < t_n$, the n-tuple $(\nu^\gamma_{t_1/\gamma}, \ldots, \nu^\gamma_{t_n/\gamma})$ converges in law for the weak topology to $(N_{t_1}\delta_{S_{t_1}}, \ldots, N_{t_n}\delta_{S_{t_n}})$ where*

(1) *$(S_t; t \geq 0)$ is a Markov jump process on $\mathcal{X}$ with initial value $S_0 = x$ and whose jumping rates $q(x, dh)$ from x to $x + h$ are given by*

$$q(x, dh) = \beta(x)\chi(x, x + h)M(x, dh),$$

where $\beta(x) = \mu(x)b(x)\mathbb{E}(\xi(x)) = \mu(x)b(x)\theta(x)/(1 - e^{-\theta(x)})$ and

$$\chi(x,y) = \sum_{n\geq 1} \frac{n\mathbb{P}(\xi(x) = n)}{\mathbb{E}(\xi(x))} u_{n,1}(x,y) = \sum_{n\geq 1} e^{-\theta(x)} \frac{\theta(x)^{n-1}}{(n-1)!} u_{n,1}(x,y). \quad (3)$$

(2) *Conditional on $(S_{t_1}, \ldots, S_{t_n}) = (x_1, \ldots, x_n)$, the N_{t_i} are independent and respectively distributed as $\xi(x_i)$.*

Therefore, in the limit of rare mutations, on the mutation timescale, the population is always monomorphic and the dominant trait of the population evolves as a jump process over the trait space, where a jump corresponds to the appearance and fixation of a mutant type. Moreover, at any time, the population size is stationary (i.e. has the stationary distribution corresponding to the dominant trait of the population). The fixation rate of a mutant is governed by the function $\chi(x,y)$, which is therefore the random analogue of the traditional invasion fitness [9], defined as the probability of invasion of a mutant type y in a resident population of type x at equilibrium. Observe that, as usual in adaptive dynamics, the fitness landscape depends on the current state of the population. Moreover, in contrast with the classical TSS [10], from a given monomorphic resident population, any mutant trait has a positive probability to invade (genetic drift). Therefore, evolution is possible in any direction of the trait space. However, selection still acts in a preferential direction, as will appear in the next section.

We refer to [2] for the proof of this result. However, this convergence is natural in view of the following interpretation of each parameter. The quantity $\beta(x)$ can be seen as the *mean mutant production rate* of a stationary x-type population (i.e. with size $\xi(x)$), and $\chi(x,y)$ is the *probability of fixation* of a single y-type mutant entering a pure x-type population with *size-biased stationary size*. The size bias comes from the fact that the mutant appears *at a birth time* in the stationary population (since the birth rate is proportional to the population size, the population size after a birth event in the stationary population is given by the size-biaised stationary population size).

4. The canonical diffusion of adaptive dynamics. Let us assume for simplicity that $\mathcal{X} = \mathbb{R}^k$. Let $\sigma(x)$ be the square root matrix of the covariance matrix of $M(x,\cdot)$. We also need to assume that the matrix $\sigma(x)$ is a Lipschitz function of x.

In order to obtain the equivalent of the canonical equation of adaptive dynamics in a finite population, we want to apply a limit of small mutation steps (weak selection) to the TSS S. To this aim, we introduce a parameter $\epsilon > 0$ and replace the mutation kernels $M(x,\cdot)$ by their image by the application $h \mapsto \epsilon h$. Time also has to be rescaled in order to obtain a non-degenerate limit. The correct time scaling is $1/\epsilon^2$, which leads to the following generator for the rescaled TSS $(S^\epsilon_t; t \geq 0)$

$$A_\epsilon \varphi(x) = \frac{1}{\epsilon^2} \int_{\mathbb{R}^k} (\varphi(x+\epsilon h) - \varphi(x)) \beta(x) \chi(x, x+\epsilon h) M(x, dh).$$

Using the assumption that the mutation kernels $M(x,\cdot)$ have 0 expectation, it is elementary to compute the limit of this expression as $\epsilon \to 0$ (for sufficiently regular φ). This limit, which takes the form of a diffusion generator, explains the following result (its full proof can be found in [2]).

THEOREM 2. *If the family $(S^\epsilon_0)_{\epsilon>0}$ has bounded first-order moments and converges in law as $\epsilon \to 0$ to a random variable Z_0, then the process S^ϵ with initial state S^ϵ_0 converges in law for the Skorohod topology on $\mathbb{D}(\mathbb{R}_+, \mathbb{R}^k)$ to the diffusion process $(Z_t; t \geq 0)$ with initial state Z_0, the unique solution to the stochastic differential equation*

$$dZ_t = \beta(Z_t)\sigma^2(Z_t) \cdot \nabla_2 \chi(Z_t, Z_t)\, dt + \sqrt{\beta(Z_t)\chi(Z_t, Z_t)}\sigma(Z_t) \cdot dB_t \tag{4}$$

where $\nabla_2 \chi$ denotes the gradient w.r.t. the second variable y of $\chi(x,y)$ and B is a standard k-dimensional Brownian motion.

This result gives the equivalent form of the canonical equation of adaptive dynamics (1) when the population is finite. It is no longer a deterministic ODE, but a diffusion process, in which the genetic drift remains present (in the form of a stochastic diffusion term), as a consequence of the population finiteness and of the limit of weak selection. The deterministic drift part of (4) is very similar to the standard canonical equation of adaptive dynamics (1), and involves in particular the gradient of the fitness function χ. The process (4) provides a diffusion model describing the evolution of the dominant trait value in a population [8, 5], grounded on a precise microscopic density-dependent modelling of the population dynamics. It also gives the precise balance between directional selection and genetic drift as a function of the individual's dynamical parameters.

5. Fixation probability near neutrality. The SDE (4) involves the fixation probability $\chi(x,x)$ and the fitness gradient with respect to the second variable $\nabla_2\chi(x,x)$. In this section, we explain how these quantities can be explicitly computed.

We need to compute the derivatives of the fixation probabilities $u_{n,m}(x,y)$ when y is close to x. Recall that the law of the two-type LBP without mutation (X,Y) used to define $u_{n,m}$ in the end of Section 2 is characterized by the birth vector B and the competition matrix C

$$B = \begin{pmatrix} b(x) \\ b(y) \end{pmatrix}, \quad C = \begin{pmatrix} c(x,x) & c(x,y) \\ c(y,x) & c(y,y) \end{pmatrix}.$$

We will say that the mutant is *neutral* if all individuals are exchangeable, i.e. when $b(y) = b(x)$ and $c(x,y) = c(y,x) = c(y,y) = c(x,x)$ (this holds in particular when $y = x$). As will appear below, using the notation $(b,c) := (b(x), c(x,x))$, it is natural to focus on deviations from the neutral case expressed as

$$B = b\mathbf{1} + \begin{pmatrix} 0 \\ \lambda \end{pmatrix}, \quad C = c\mathbf{1} - \begin{pmatrix} 0 & 0 \\ \delta & \delta \end{pmatrix} + \begin{pmatrix} 0 & \alpha \\ 0 & \alpha \end{pmatrix} - \begin{pmatrix} 0 & \varepsilon \\ \varepsilon & 0 \end{pmatrix}.$$

In words, deviations from the neutral case are a linear combination of four *fundamental selection coefficients* λ, δ, α, ε, that are chosen to be positive when they confer an advantage to the mutant. It is convenient to assess these deviations to the neutral case in terms of

1. **fertility** (λ): positive λ means increased mutant birth rate;
2. **defence** capacity (δ): positive δ means reduced competition sensitivity of mutant individuals w.r.t. the total population size;
3. **aggressiveness** (α): positive α means raised competition pressure exerted from any mutant individual onto the rest of the population;
4. **isolation** (ε): positive ε means lighter cross-competition between the two different types, that would lead, if harsher, to a greater probability of exclusion of the less abundant one.

Under neutrality, an elementary martingale argument shows that the fixation probability $u_{n,m}$, further denoted u, equals the initial mutant frequency $p := m/(m+n)$. This implies in particular that

$$\chi(x,x) = \frac{e^{-\theta(x)} - 1 + \theta(x)}{\theta(x)^2}. \tag{5}$$

The following theorem unveils the dependence of u upon λ, δ, α, ε, when they slightly deviate from 0, and explains why these four selection coefficients provide a natural basis to decompose the gradient of the fixation probability.

THEOREM 3. *As a function of the multidimensional selection coefficient $s = (\lambda, \delta, \alpha, \varepsilon)$, the probability u is differentiable, and in a neighbourhood of $s = 0$ (selective neutrality),*

$$u = p + v'.s + o(s), \tag{6}$$

where the selection gradient $v = (v^\lambda, v^\delta, v^\alpha, v^\varepsilon)$ *can be expressed as*

$$v^\iota_{n,m} = p\,(1-p) g^\iota_{n+m}, \qquad \iota \neq \varepsilon,$$
$$v^\varepsilon_{n,m} = p\,(1-p)\,(1-2p)\; g^\varepsilon_{n+m}$$

where the g*'s depend solely on the resident's characteristics* b, c*, and on the total initial population size* $n+m$*. They are called the invasibility coefficients.*

The invasibility coefficients of a pure resident population are interesting to study, as they provide insights about how the fixation probability deviates from p as the selection coefficients of the mutant deviate from 0. They provide information about the *robustness* of the resident population, i.e. its resistance to mutant invasions. In the simplest case where mutations in the parameter space are isotropic, the biggest invasibility coefficient gives the direction of the parameter space where a mutant is more likely to invade. More generally, when there are correlations between mutations in the parameter space, the likeliest direction of evolution in the trait space is given by the deterministic coefficient of the canonical diffusion (4), in which the fitness gradient is given by

$$\nabla_2 \chi(x,x) = a_\lambda(x) \nabla b(x) - a_\delta(x) \nabla_1 c(x,x) + a_\alpha(x) \nabla_2 c(x,x), \tag{7}$$

where, for $\iota = \lambda, \delta, \alpha$,

$$a_\iota(x) = e^{-\theta(x)} \sum_{n=1}^{\infty} \frac{n g^\iota_{n+1}(x) \theta(x)^{n-1}}{(n+1)^2 (n-1)!}.$$

It is possible to obtain explicit expressions for the invasibility coefficients g^ι as series. We refer to [2] for the exact expressions. In particular, these expressions yield that $a_\iota(x) = \hat{a}_\iota(\theta(x))/c(x,x)$ for some function $\hat{a}_\iota$. Moreover, they allow one to compute numerically the invasibility coefficients, and therefore the quantities $\hat{a}_\iota$ for $\iota = \lambda, \delta, \alpha$ as functions of the parameter $\theta(x)$. These numerical results can be used to make simulations of the canonical diffusion of adaptive dynamics in various ecological examples. In particular, in contrast with the classical canonical equation of adaptive dynamics, the presence of a genetic drift can induce the evolutionary dynamics to drift away from evolutionary stable strategies, where the fitness gradient is zero. When the fitness gradient admits several zeros, it can visit several basins of attraction on various timescales.

This numerical study is a work in progress, that is quite delicate because the series involved in the computation of g^ι are slowly converging, with first terms that grow exponentially fast with θ. We shall give here some preliminary results. Fig. 1 shows the functions $\hat{a}_\iota$ for $\iota = \lambda, \delta, \alpha$. Several comments can be made from this figure. First, for any $\theta > 0$, $\hat{a}_\delta(\theta) > \hat{a}_\alpha(\theta) > \hat{a}_\lambda(\theta)$. This means that, for equal mutation steps in the parameter space, a mutation is always more advantageous in the direction δ than in the direction α, which is itself more advantageous than in the direction λ. In other words, in a given population, a better defence capacity is more beneficial than a harsher aggressiveness, which is more beneficial than a greater fertility.

Moreover, as θ goes to infinity, these functions have different asymptotic behaviours. $\hat{a}_\delta$ seems to converge to $1/2$, whereas $\hat{a}_\lambda(\theta)$ and $\hat{a}_\alpha(\theta)$ are both equivalent to $1/2\theta$ (see Fig. 2). However, this does not mean that mutations are more likely to become fixed in the δ direction when θ is large, because the fact that $b = \theta c$ must be taken into account.

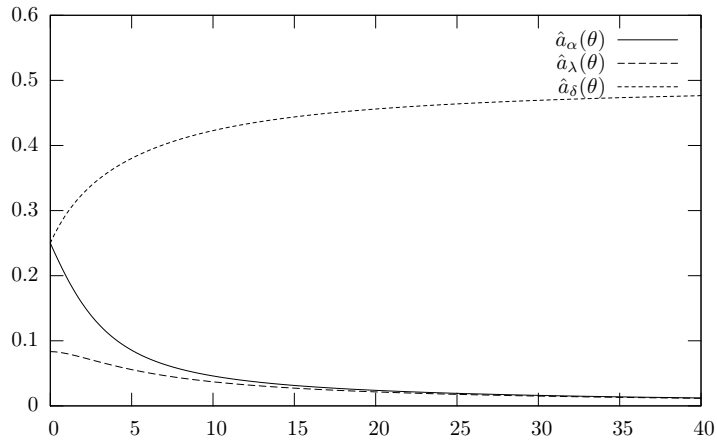


Fig. 1. The functions $\hat{a}_\lambda$, $\hat{a}_\delta$ and $\hat{a}_\alpha$ as functions of θ

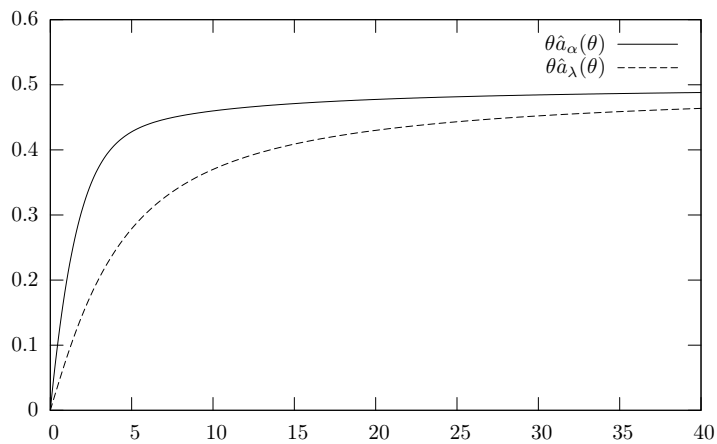


Fig. 2. The functions $\theta \mapsto \theta\hat{a}_\lambda(\theta)$ and $\theta \mapsto \theta\hat{a}_\alpha(\theta)$

The precise balance between the functions a_ι in (7) when the population is large can be computed as follows: introduce a parameter K in the microscopic model by dividing $c(\cdot,\cdot)$ by K and consider the limit $K \to +\infty$ in the canonical diffusion. Denoting by $\chi_K(\cdot,\cdot)$ the fitness function obtained this way and using the asymptotic behaviours given above, one gets that

$$\lim_{K\to+\infty} \nabla_2\chi_K(x,x) = \frac{1}{2b(x)}(\nabla b(x) - \theta(x)\nabla_1 c(x,x)).$$

Moreover, by (5), $\chi_K(x,x)$ converges to 0 when $K \to \infty$. Now, as proved in [1], the fitness function of the canonical equation of adaptive dynamics (1) is given by $f(x,y) = b(y) - c(y,x)\theta(x)$. Therefore, as $K \to +\infty$, the canonical diffusion converges to a deterministic ODE which is precisely the canonical equation of adaptive dynamics. This gives a new justification of this equation, and allows one to study the fluctuations around the

canonical equations when K is large but finite. In particular, our diffusion equation with small diffusion term enters the framework of Freidlin-Wentzell's theory [4], which can be applied to predict the long time behaviour of the diffusion, and its chain of visit of basins of attractions when K is large. This kind of information is biologically very relevant, since it allows one to predict in which order all the evolutionary stable strategies will be visited by the population and on which timescale.

References

[1] N. Champagnat, *A microscopic interpretation for adaptive dynamics trait substitution sequence models*, Stoch. Proc. Appl. 116 (2006), 1127–1160.

[2] N. Champagnat and A. Lambert, *Evolution of discrete populations and the canonical diffusion of adaptive dynamics*, Ann. Appl. Prob. 17 (2007), 102–155.

[3] U. Dieckmann and R. Law, *The dynamical theory of coevolution: A derivation from stochastic ecological processes*, J. Math. Biol. 34 (1996), 579–612.

[4] M. I. Freidlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems*, Springer-Verlag, Berlin, 1984.

[5] T. F. Hansen, *Stabilizing selection and the comparative analysis of adaptation*, Evolution 51 (1997), 1341–1351.

[6] J. Hofbauer and R. Sigmund, *Adaptive dynamics and evolutionary stability*, Appl. Math. Letters 3 (1990), 75–79.

[7] A. Lambert, *The branching process with logistic growth*, Ann. Appl. Prob. 15 (2005), 1506–1535.

[8] R. Lande, *Natural selection and random genetic drift in phenotypic evolution*, Evolution 30 (1976), 314–334.

[9] J. A. J. Metz, R. M. Nisbet and S. A. H. Geritz, *How should we define 'fitness' for general ecological scenarios?*, Trends in Ecology and Evolution 7 (1992), 198–202.

[10] J. A. J. Metz, S. A. H. Geritz, G. Meszéna, F. A. J. Jacobs and J. S. van Heerwaarden, *Adaptive Dynamics, a geometrical study of the consequences of nearly faithful reproduction*, in: Stochastic and Spatial Structures of Dynamical Systems, S. J. van Strien and S. M. Verduyn Lunel (eds.), North-Holland, Amsterdam, 1996, 183–231.

STOCHASTIC MODELS IN BIOLOGICAL SCIENCES
BANACH CENTER PUBLICATIONS, VOLUME 80
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WARSZAWA 2008

STOCHASTIC STABILITY IN SPATIAL GAMES

JACEK MIĘKISZ

Institute of Applied Mathematics and Mechanics, University of Warsaw
Banacha 2, 02-097 Warsaw, Poland
E-mail: miekisz@mimuw.edu.pl

Abstract. We compare two concepts of stochastic stability in spatial games. The classical approach to stochastic stability, introduced by Foster and Young [8], involves single configurations in the zero-noise limit. Ensemble stability discussed in [17] refers to ensembles of configurations in the limit of an infinite number of players. The above two limits may not commute. We will discuss reasons of such behaviour. We review some results concerning the effect of the number of players and the noise level on the long-run behaviour of stochastic dynamics of spatial games.

1. Introduction. Many evolutionary and ecological processes can be modeled as systems of interacting players [16, 33, 32, 27, 12, 35, 5, 13, 24, 25]. In such models, players have at their disposal certain strategies and their payoffs depend on strategies chosen both by them and by their opponents. The central concept in game theory is that of a *Nash equilibrium*. It is an assignment of strategies to players such that no player, for fixed strategies of his opponents, has an incentive to deviate from his current strategy; no change can increase his payoff. One of the fundamental problems in game theory is the equilibrium selection in games with multiple Nash equilibria. One of the selection methods is to construct an appropriate dynamical system where in the long run only one equilibrium is played with a high frequency.

Here we will discuss spatial games with certain local interactions. In such games, players are located on vertices of certain graphs and they interact only with their neighbours; see for example [21, 22, 23, 1, 6, 35, 7, 15, 30, 11] and a recent review [31] and references therein. In discrete moments of time, players adapt to their neighbors by choosing with a high probability the strategy which is the best response, i.e. the one which maximizes

2000 *Mathematics Subject Classification*: 91A10, 91A22.

Key words and phrases: evolutionary game theory, Nash equilibria, spatial games, equilibrium selection, stochastic stability.

I would like to thank the Ministry of Science and Higher Education for a financial support under the grant N201 023 31/2069.

The paper is in final form and no version of it will be published elsewhere.

 © Instytut Matematyczny PAN, 2008

the sum of the payoffs obtained in individual games and with a small probability they make mistakes.

To describe the long-run behavior of such stochastic dynamics, Foster and Young [8] introduced a concept of stochastic stability. A configuration of a system—an assignment of stategies to players—is *stochastically stable* if it has a positive probability in the stationary distribution of the above dynamics in the limit of zero noise. It means that in the long run we observe it with a positive frequency. In classical applications of stochastic stability for equilibrium selection [14, 26, 32, 34, 35], the population size was fixed and the noise level was taken to zero. However, we may also consider models with a fixed noise level in the limit of an infinite number of players [27, 17, 18, 19, 20, 28]. Now, for any arbitrarily low but fixed noise, if the number of players is big enough, then the probability of any individual configuration is practically zero. It means that for a big number of players, to observe a stochastically stable configuration we must assume that players make mistakes with extremely small probabilities. On the other hand, it may happen that in the long run, for a low but fixed noise and sufficiently big number of players, the stationary distribution is highly concentrated on an ensemble consisting of one Nash configuration and its small perturbations, i.e. configurations where most players play the same strategy. We will call such configurations *ensemble stable* [17].

In Section 2, we formally introduce spatial games and two concepts of stochastic stability. In Section 3, we discuss some other results and state an open problem.

2. Spatial games and stochastic stability. Let Λ be a finite subset of the simple lattice $\mathbf{Z}^d$ (we may think about a square centered at the origin of the lattice). Every site of Λ is occupied by one player who has at his disposal one of k different pure strategies. Let S be a set of pure strategies, then $\Omega_\Lambda = S^\Lambda$ is the set of all possible configurations of players, that is all possible assignments of pure strategies to individual players. For every $i \in \Lambda$, X_i is the strategy of the i-th player in the configuration $X \in \Omega_\Lambda$ and X_{-i} denotes strategies of all remaining players; X therefore can be represented as the pair (X_i, X_{-i}). $U : S \times S \to R$ is a matrix of payoffs of our stage game. $U(A, B), A, B \in S$ is the payoff of the first (row) player playing the strategy A when the second one (a column player) is playing B. We will consider here only symmetric games so the payoff of the second player is given by $U(B, A)$ (the payoff matrix of the second player is the transpose of the payoff matrix U of the first one). Every player interacts only with his neighbours and his payoff is the sum of the payoffs resulting from individual games. We assume that he has to use the same strategy for all neighbours. Let N_i denote the neighbourhood of the i-th player. For the nearest-neighbor interaction we have $N_i = \{j; |j - i| = 1\}$, where $|i - j|$ is the distance between i and j. For $X \in \Omega_\Lambda$ we denote by $\nu_i(X)$ the payoff of the i-th player in the configuration X:

$$\nu_i(X) = \sum_{j \in N_i} U(X_i, X_j) \tag{1}$$

Definition 1. $X \in \Omega_\Lambda$ is a *Nash configuration* if for every $i \in \Lambda$ and $A \in S$

$$\nu_i(X_i, X_{-i}) \geq \nu_i(A, X_{-i}).$$

We describe now the deterministic dynamics of the *best-response rule*. Namely, at each discrete moment of time $t = 1, 2, ...$, a randomly chosen player may update his strategy. He simply adopts the strategy, X_i^t, which gives him the maximal total payoff $\nu_i(X_i^t, X_{-i}^{t-1})$ for given X_{-i}^{t-1}, a configuration of strategies of remaining players at time $t-1$.

Now we allow players to make mistakes with a small probability, that is to say they may not choose the best response. A probability of making a mistake may depend on the state of the system (a configuration of strategies of neighboring players). We will assume that this probability is a decreasing function of the payoff lost as a result of a mistake [1, 2]. In the *log-linear rule*, the probability of chosing by the i-th player the strategy X_i^t at time t is given by the following conditional probability:

$$p_i^\epsilon(X_i^t|X_{-i}^{t-1}) = \frac{e^{(1/\epsilon)\nu_i(X_i^t, X_{-i}^{t-1})}}{\sum_{A\in S} e^{(1/\epsilon)\nu_i(A, X_{-i}^{t-1})}}, \tag{2}$$

where $\epsilon > 0$ measures the noise level.

Let us observe that if $\epsilon \to 0$, p_i^ϵ converges to the best-response rule. Our stochastic dynamics is an example of an ergodic Markov chain with $|S^\Lambda|$ states. Therefore, it has a unique stationary distribution which we denote by μ_Λ^ϵ.

The following definition was first introduced by Foster and Young [8]:

DEFINITION 2. $X \in \Omega_\Lambda$ is *stochastically stable* if $\lim_{\epsilon\to 0} \mu_\Lambda^\epsilon(X) > 0$.

If X is stochastically stable, then the frequency of visiting X converges to a positive number along any time trajectory almost surely. It means that in the long run we observe X with a positive frequency. In most models it is usually equal to 1. In examples below, we consider symmetric games with symmetric Nash equilibria and homogeneous Nash configurations. By stochastic stability of a strategy or a Nash equilibrium we mean stochastic stability of the corresponding Nash configuration.

EXAMPLE 1. Players are located on a finite subset Λ of $\mathbf{Z}$ (with periodic boundary conditions) and interact with their two nearest neighbors. They have at their disposal two strategies: A and B. The payoffs are given by the following symmetric matrix:

$$U = \begin{array}{c|cc} & A & B \\ A & 5 & 0 \\ B & 3 & 3 \end{array}$$

It is easy to see that the above stage game has two Nash equilibria: (A, A) and (B, B), and the corresponding spatial game has two homogeneous Nash configurations: X^A and X^B, in which all individuals play the same strategy, A or B respectively.

One can show that X^B is stochastically stable [1, 34]. The reason for such a long-run behavior is that the population needs many mistakes (actually the number of required mistakes is proportional to the population size) to move from X^B to X^A. On the contrary, one needs only one mistake to move from X^A to X^B; one A-player makes a mistake and switches to B and then both his neighbours may switch to B by the noise-free dynamics of the best-response rule. We see that no more mistakes are required for B strategy to invade the whole population. One can make this argument rigorous using a tree representation of stationary distributions of ergodic Markov chains [9, 14, 34, 26, 32, 35, 6, 7, 20].

In general, we have to investigate all absorbing states of the noise-free dynamics. When we start with any non-absorbing state, then after a finite number of steps of the noise-free best-response dynamics we arrive at one of the absorbing states and then stay there forever. It follows from the tree representation of the stationary distribution that any state different from absorbing states has zero probability in the stationary distribution in the zero-noise limit.

Stochastically stable states are those absorbing states from which it is relatively more difficult to get out than from other absorbing states [6, 7].

The situation gets more complex when for a fixed noise level we are approaching the infinite population limit. It may happen then that in the long run one sees configurations from which it is relatively easier for a player to switch to a different strategy than from other absorbing configurations. This is the case in the following example.

EXAMPLE 2. Players are located on a finite subset Λ of $\mathbf{Z}^2$ (with periodic boundary conditions) and interact with their four nearest neighbours. They have at their disposal three pure strategies: A, B, and C. The payoffs are given by the following symmetric matrix:

$$U = \begin{array}{c|ccc} & A & B & C \\ A & 2 & 0 & 1 \\ B & 0 & 3+\alpha & 0 \\ C & 1 & 0 & 3, \end{array}$$

where $\alpha \geq 0$.

Our game has three Nash equilibria: (A, A), (B, B), and (C, C), and the corresponding spatial game has three homogeneous Nash configurations: X^A, X^B, and X^C.

The unique stationary distribution of the log-linear dynamics of the above game can be explicitly constructed [35] and is given by

$$\mu^{\epsilon}_{\Lambda}(X) = \frac{e^{(1/\epsilon)\sum_{(i,j)\in\Lambda} U(X_i,X_j)}}{\sum_{Z\in\Omega_\Lambda} e^{(1/\epsilon)\sum_{(i,j)\in\Lambda} U(Z_i,Z_j)}}. \tag{3}$$

Let us start our discussion with the case of $\alpha = 0$. It follows from (3) that

$$\lim_{\epsilon\to 0} \mu^{\epsilon}_{\Lambda}(X^k) = 1/2,$$

for $k = B, C$ so B and C are stochastically stable.

Let us investigate the long-run behavior of our system for large Λ, that is for a big number of players. Observe that

$$\lim_{\Lambda\to\mathbf{Z}^2} \mu^{\epsilon}_{\Lambda}(X) = 0$$

for every $X \in \Omega = S^{\mathbf{Z}^2}$.

Hence for large Λ and $\epsilon > 0$ we may only observe, with reasonable positive frequencies, ensembles of configurations and not particular configurations. We will be interested in ensembles which consist of a Nash configuration and its small perturbations, that is configurations, where most players adopt the same strategy.

We perform first the limit $\Lambda \to \mathbf{Z}^2$ and obtain an infinite-population state, the so called Gibbs state [10, 29],

$$\mu^\epsilon = \lim_{\Lambda \to \mathbf{Z}^2} \mu^\epsilon_\Lambda. \tag{4}$$

We use techniques of statistical mechanics and the concept of dominant ground states [3, 4], modify proofs in [17] and get

$$\mu^\epsilon(X_i \neq C) < \gamma(\epsilon) \tag{5}$$

for any $i \in \mathbf{Z}^2$ and $\gamma(\epsilon) \to 0$ as $\epsilon \to 0$.

Let us sketch briefly the reason of such a result. While it is true that both X^B and X^C have the same maximal total payoff (and therefore are stochastically stable), the X^C Nash configuration has lower-cost excitations. Namely, one player can change its strategy and switch to A and the total payoff will decrease by 16 units. Players in the X^B Nash configuration can switch either to A or B to decrease the total payoff by 24. Now, the probability of the occurrence of any configuration in the Gibbs state (which is the stationary distribution of our stochastic dynamics) depends on the total payoff in an exponential way. One can prove that the probability of the ensemble consisting of the X^C Nash configuration and configurations which are different from it at few sites only is much bigger than the probability of the analogous X^B-ensemble. On the other hand, configurations which are outside X^B and X^C-ensembles appear with exponentially small probabilities. It means that for large enough systems (and small but not extremely small ϵ) we observe in the stationary distribution the X^C Nash configuration with perhaps few different strategies. The above argument was made into a rigorous proof for an infinite system of the closely related lattice-gas model (the Blume-Capel model) of interacting particles by Bricmont and Slawny in [3].

We see that for any low but fixed ϵ, if the number of players is big enough, then in the long run, almost all players use C strategy. On the other hand, if for any fixed number of players, ϵ is lowered substantially, then B and C appear with frequencies close to 1/2.

The above example motivates following definitions [17].

DEFINITION 3. $X \in \Omega_\Lambda$ is *γ-ensemble stable* if $\mu^\epsilon_\Lambda(Y \in \Omega_\Lambda; Y_i \neq X_i) < \gamma$ for any $i \in \Lambda$ if $\Lambda \supset \Lambda(\gamma)$ for some $\Lambda(\gamma)$.

DEFINITION 4. $X \in \Omega_\Lambda$ is *low-noise ensemble stable* if for every $\gamma > 0$ there exists $\epsilon(\gamma)$ such that if $\epsilon < \epsilon(\gamma)$, then X is γ-ensemble stable.

The following theorem is a simple consequence of (5).

THEOREM 1. *If $\alpha = 0$, then X^C is low-noise ensemble stable.*

If X is γ-ensemble stable with γ close to zero, then the ensemble consisting of X and configurations which are different from X at at most few sites has the probability close to one in the stationary distribution. It does not follow, however, that X is necessarily low-noise ensemble or stochastically stable as it happens in our example in the case of $\alpha > 0$.

For $\alpha = 0$, both X^B and X^C have the same total payoff. X^C has lower-cost fluctuations and therefore it is low-noise ensemble stable. For $\alpha > 0$, X^C has a smaller

total payoff but nevertheless one can prove [17] that in the long run C is played with a frequency close to 1 if the noise level is low but not extremely low.

THEOREM 2. *For every $\gamma > 0$, there exist $\alpha(\gamma)$ and $\epsilon(\gamma)$ such that for every $0 < \alpha < \alpha(\gamma)$, there exists $\epsilon(\alpha)$ such that for $\epsilon(\alpha) < \epsilon < \epsilon(\gamma)$, X^C is γ-ensemble stable, and for $0 < \epsilon < \epsilon(\alpha)$, X^B is γ-ensemble stable.*

Observe that for $\alpha = 0$, both X^B and X^C are stochastically stable (they appear with the frequency $1/2$ in the limit of zero noise) but X^C is low-noise ensemble stable. For small $\alpha > 0$, X^B is both stochastically (it appears with the frequency 1 in the limit of zero noise) and low-noise ensemble stable. However, for an intermediate noise $\epsilon(\alpha) < \epsilon < \epsilon(\gamma)$, if the number of players is large enough, then in the long run, almost all players use the strategy C (X^C is ensemble stable). If we lower ϵ below $\epsilon(\alpha)$, then almost all players start to use the strategy B. $\epsilon = \epsilon(\alpha)$ is the line of the first-order phase transition. In the infinite population limit there exist two stationary distributions on this line. We may say that at $\epsilon = \epsilon(\alpha)$, the population of players undergoes a sharp *equilibrium transition* from C to B-behaviour.

3. Discussion. We have discussed two concepts of stochastic stability in spatial games. Classical stochastic stability involves single configurations in the limit of zero noise, ensemble stability refers to ensembles of configurations in the infinite population limit. We have shown that the above two limits may not commute. Therefore, to study the long-run behavior of real populations, we should estimate the relevant parameters to be sure what limiting procedures are appropriate in specific examples.

Similar ideas were recently discussed in two-player games in well-mixed populations [28]. States of such systems are described by strategy frequencies. One takes first the limit of an infinite number of players. Stationary distributions are then measures concentrated on single frequencies. It was shown in [28] that a stationary frequency may change discontinously at a certain noise level.

Let us end our paper by presenting an open problem.

OPEN PROBLEM. Construct a stochastic evolutionary spatial game with a unique stationary distribution μ_Λ^ϵ such that

$$\lim_{\epsilon\to 0}\lim_{\Lambda\to\mathbf{Z}^2}\mu_\Lambda^\epsilon(X) = 1,$$
$$\lim_{\Lambda\to\mathbf{Z}^2}\lim_{\epsilon\to 0}\mu_\Lambda^\epsilon(Y) = 1,$$

where $X, Y \in \Omega = Z^d$, $X \neq Y$.

References

[1] L. E. Blume, *The statistical mechanics of strategic interaction,* Games Econ. Behav. 5 (1993), 387–424.

[2] L. E. Blume, *How noise matters,* Games Econ. Behav. 44 (2003), 251–271.

[3] J. Bricmont and J. Slawny, *First order phase transitions and perturbation theory,* in: Statistical Mechanics and Field Theory: Mathematical Aspects, Lecture Notes in Physics 257, Springer-Verlag, 1986.

[4] J. Bricmont and J. Slawny, *Phase transitions in systems with a finite number of dominant ground states,* J. Stat. Phys. 54 (1989), 89–161.

[5] R. Cressman, *Evolutionary Dynamics and Extensive Form Games,* MIT Press, Cambridge, USA, 2003.

[6] G. Ellison, *Learning, local interaction, and coordination,* Econometrica 61 (1993), 1047–1071.

[7] G. Ellison, *Basins of attraction, long-run stochastic stability, and the speed of step-by-step evolution,* Rev. Econ. Stud. 67 (2000), 17–45.

[8] D. Foster and P. H. Young, *Stochastic evolutionary game dynamic,* Theor. Pop. Biol. 38 (1990), 219–232.

[9] M. I. Freidlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems,* Springer Verlag, New York, 1984.

[10] H. O. Georgii, *Gibbs Measures and Phase Transitions,* Walter de Gruyter, Berlin, 1988.

[11] C. Hauert and M. Doebeli, *Spatial structure often inhibits the evolution of cooperation in the snowdrift game,* Nature 428, (2004), 643–646.

[12] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics,* Cambridge University Press, Cambridge, 1998.

[13] J. Hofbauer and K. Sigmund, *Evolutionary game dynamics,* Bull. Am. Math. Soc. 40 (2003), 479–519.

[14] M. Kandori, G. J. Mailath, and R. Rob, *Learning, mutation, and long-run equilibria in games,* Econometrica 61 (1993), 29–56.

[15] K. Lindgren and M. G. Nordahl, *Evolutionary dynamics of spatial games,* Physica D 75 (1994), 292–309.

[16] J. Maynard Smith, *Evolution and the Theory of Games,* Cambridge University Press, Cambridge, 1982.

[17] J. Miękisz, *Statistical mechanics of spatial evolutionary games,* Phys. A: Math. Gen. 37 (2004), 9891–9906.

[18] J. Miękisz *Stochastic stability in spatial games,* J. Stat. Phys. 117 (2004), 99–110.

[19] J. Miękisz, *Stochastic stability in spatial three-player games,* Physica A 343 (2004), 175–184.

[20] J. Miękisz, *Equilibrium selection in evolutionary games with random matching of players,* J. Theor. Biol. 232 (2005), 47–53.

[21] M. A. Nowak and R. M. May, *Evolutionary games and spatial chaos,* Nature 359(1992), 826–829.

[22] M. A. Nowak and R. M. May, *The spatial dilemmas of evolution,* Int. J. Bifurcation Chaos 3 (1993), 35–78.

[23] M. A. Nowak, S. Bonhoeffer, and R. M. May, *More spatial games,* Int. J. Bifurcation Chaos 4 (1994), 33–56.

[24] M. A. Nowak and K. Sigmund, *Evolutionary dynamics of biological games,* Science 303 (2004), 793–799.

[25] M. A. Nowak, *Evolutionary Dynamics,* Harvard University Press, Cambridge, USA, 2006.

[26] A. Robson and F. Vega-Redondo, *Efficient equilibrium selection in evolutionary games with random matching,* J. Econ. Theory 70 (1996), 65–92.

[27] L. Samuelson, *Evolutionary Games and Equilibrium Selection,* MIT Press, Cambridge, USA, 1997.

[28] W. H. Sandholm, *Simple formulas for stationary distributions and stochastically stable states,* Games Econ. Behav. 59 (2007), 154–162.

[29] Y. G. Sinai, *Theory of Phase Transitions: Rigorous Results,* Pergamon, Oxford, 1982.

[30] G. Szabó, J. Vukov, and A. Szolnoki, *Phase diagrams for an evolutionary prisoner's dilemma game on two-dimensional lattices,* Phys. Rev. E 72 (2005), 047107.

[31] G. Szabó and G. Fáth, *Evolutionary games on graphs,* Physics Reports 446 (2007), 97–216.

[32] F. Vega-Redondo, *Evolution, Games, and Economic Behaviour,* Oxford University Press, Oxford, 1996.

[33] J. Weibull, *Evolutionary Game Theory,* MIT Press, Cambridge, USA, 1995.

[34] H. P. Young, *The evolution of conventions,* Econometrica 61 (1993), 57–84.

[35] H. P. Young, *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions,* Princeton University Press, Princeton, 1998.

STOCHASTIC MODELS IN BIOLOGICAL SCIENCES
BANACH CENTER PUBLICATIONS, VOLUME 80
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WARSZAWA 2008

SPATIAL STOCHASTIC PREDATOR-PREY MODELS

MAURO MOBILIA,[1,2] IVAN T. GEORGIEV,[2] and UWE C. TÄUBER[2]

[1] *Arnold Sommerfeld Center for Theoretical Physics and Center for NanoScience*
Department of Physics, Ludwig-Maximilians-Universität München
D-80333 München, Germany
[2] *Department of Physics and Center for Stochastic Processes in Science and Engineering*
Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061-0435, U.S.A.
E-mail: mauro.mobilia@physik.lmu.de, ivantgeorgiev@gmail.com, tauber@vt.edu

Abstract. We consider a broad class of stochastic lattice predator-prey models whose main features are overviewed. In particular, this article aims at drawing a picture of the influence of spatial fluctuations, which are not accounted for by the deterministic rate equations, on the properties of the stochastic models. Here, we outline the robust scenario obeyed by most of the lattice predator-prey models with an interaction *à la Lotka-Volterra*. We also show how a drastically different behavior can emerge as the result of a subtle interplay between long-range interactions and a nearest-neighbor exchange process.

1. Introduction. Understanding biodiversity is a central challenge in modern evolutionary and theoretical biology [May]. In this realm, studying population dynamic models to understand and identify the mechanisms allowing for coevolution of competing species is an important topic. The latter is classically addressed by considering deterministic nonlinear differential equations. Within this approach, the set of equations devised many decades ago by Lotka and Volterra [Lot] is certainly a paradigm. These authors considered two coupled nonlinear differential equations mimicking the evolution of a two-species competing system. Hence, within their model, Lotka and Volterra demonstrated that the coexistence of the species always occurs and that the densities of the populations regularly oscillate in time. In spite of their popularity, the Lotka-Volterra equations have often been criticized as being biologically unrealistic and mathematically unstable [May]. Actually,

2000 *Mathematics Subject Classification*: Primary 92D25; Secondary 82C26, 82C31.
Key words and phrases: stochastic predator-prey models.
This research was in part supported by the US NSF grant DMR-0308548 and (for M.M.) by Swiss NSF fellowship 81EL-68473. M.M. gratefully acknowledges the support of the Alexander von Humboldt Foundation (fellowship IV-SCZ/1119205 STP).
The paper is in final form and no version of it will be published elsewhere.

 © Instytut Matematyczny PAN, 2008

to gain some more realistic and fundamental understanding on population coevolution and biodiversity, it is important to take into account discrete spatial degrees of freedom by going beyond the deterministic picture. In this context, various stochastic predator-prey models have recently been investigated. In this article, based on Refs. [Mob1,Mob2] where more details can be found, we aim at drawing a brief overview of the robust properties of these lattice systems with Lotka-Volterra interactions. Also, by considering a simple ecological model, we will show how the subtle interplay between the degree of mixing of a stochastic system and long-range interaction can give rise to surprising features.

2. Deterministic rate equations. It is useful to first review some results of the deterministic approach for the two-species system, where predators (A) and prey (B) interact according to the following reactions: $A \to \oslash$ (death rate μ), $A + B \to A + A$ (predation rate λ) and $A \to A + A$ (reproduction rate σ). Neglecting any spatial variations of the densities $a(t)$ and $b(t)$ of predators and prey, respectively, one obtains the classical Lotka-Volterra equations:

$$\dot{a}(t) = \lambda a(t)b(t) - \mu a(t); \quad \dot{b} = \sigma b(t) - \lambda a(t)b(t). \tag{1}$$

A linear stability analysis of these equations show that the densities oscillate around the center (neutrally stable) fixed point $(a_c, b_c) = (\sigma/\lambda, \mu/\lambda)$ with a characteristic frequency $f = \sqrt{\mu\sigma}/2\pi$. In addition, the existence of a conserved first integral $K(t) = \lambda[a(t) + b(t)] - \sigma \ln a(t) - \mu \ln b(t)$ implies oscillatory kinetics and coexistence in the whole phase portrait, which is characterized by cyclic trajectories.

As the above deterministic cycles are unstable against any perturbations and the solutions of Eqs. (1) display amplitudes depending on the initial conditions, which is clearly an unrealistic feature, the rate equations are often rendered more realistic by including growth-limiting terms. For instance, by assuming that the prey carrying capacity is ρ, one is led to the following equation for species B:

$$\dot{b}(t) = \sigma b(t)[1 - b(t)/\rho] - \lambda a(t)b(t). \tag{2}$$

In this case, the new rate equations exhibit an extinction threshold and have two stable fixed points: (i) $(0, \rho)$, corresponding to a system full of prey and extinction of the predators, which is a stable node when $\lambda < \lambda_c = \mu/\rho$ (extinction threshold); and (ii) $([1 - \mu/\lambda\rho]\sigma/\lambda, \mu/\lambda)$, which is associated with the coexistence phase and is either a stable node (near the extinction threshold) or a focus (deep in the coexistence phase). The existence of a Lyapunov function ensures that these fixed points are actually *globally stable* [May, Mob1, Mob2].

3. The stochastic lattice Lotka-Volterra model (and its variants). To probe the deterministic approach and gain some understanding on the role of spatial fluctuations and correlations, we have studied a stochastic lattice Lotka-Volterra model (SLLVM). The latter is formulated in the natural language of reaction-diffusion systems and is defined on a $d-$dimensional lattice, on which the above Lotka-Volterra reactions are implemented. In our modeling, we mimic spatial limitation of the resources by assuming that each lattice site can be at most occupied by one predator or one prey (site restriction). Monte Carlo simulations of the SLLVM show that, in dimensions $d > 1$, the phase portrait displays

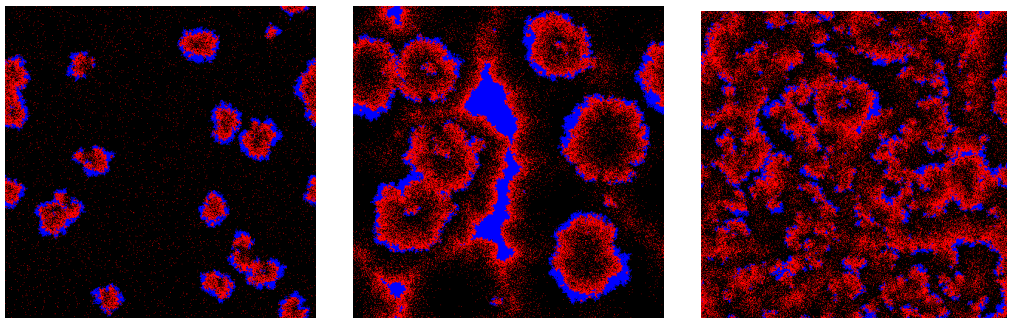


Fig. 1 (color online). Reproduced from [Mob2]. Snapshots of the time evolution (time increases from left to right) of the SLLVM on a 512×512 lattice, with rates $\sigma = 4.0, \mu = 0.1, \lambda = 2.2$. The red (light grey), blue (grey) and dark dots respectively represent the prey, predators and empty sites. Initially the system is homogeneous with densities $a(0) = b(0) = 1/3$.

qualitatively the behavior predicted by the rate equations with growth-limiting term (see Fig. 1 of [Mob2]). In fact, it is found that above some critical threshold both predators and prey coexist. The related fixed point is either a node or, for high predation rate, a focus which is associated with spiraling flows. In the latter case, as shown in Fig. (1), rich spatiotemporal patterns of persistent predator-prey "pursuit and evasion" waves (see e.g. Murray's book [May]) develop and translate into erratic damped population density oscillations [Mob2]. Obviously these features are not captured by the rate equations (1,2). In *finite* systems, the quasiperiodic fluctuations appear on a global scale, with amplitudes vanishing with the system size. We have analyzed the spatial structure and the time evolution of the above complex patterns by computing the stationary two-point correlation functions and the power spectrum [Mob2]. This has led us to an understanding of the typical size of the clusters reported on Fig. (1, rightmost) and has shown that the characteristic frequency of the SLLVM is markedly smaller than the ones predicted by the deterministic equations [Mob2]. A completely different picture emerges when the reactive fixed point is a node, just above the extinction threshold (λ_c) for the predators. Here, one observes clouds of predators effectively diffusing in a sea of prey (see Fig. 3 of [Mob2]). At the critical value, the system undergoes an *active-to-absorbing phase transition*. Various critical exponents were computed (both in 2D and 3D) and found to be in agreement with those of the *directed percolation* (DP) [Jans]. By deriving an equivalent field-theory action from the master equation of the SLLVM and by mapping it (near λ_c) onto Reggeon field-theory, we provided analytic arguments supporting the numerical indications that such a transition is in the DP universality class [Mob2]. Similar DP exponents were found numerically in many other stochastic predator-prey models (see e.g. [Ant]). As a further probe of the robustness of the above scenario, we have considered different variants of the stochastic model, namely the SLVVM supplemented with (i) diffusing predators and prey (with same diffusivities); (ii) biased diffusion of both species; (iii) predators "following" the prey, with biased hoping rates towards neighboring sites occupied by prey. In all these cases, the above features were reproduced. Let us also note that for SLLVM without site restriction it was shown that predators and prey always coexist in 1D and 2D [Wash].

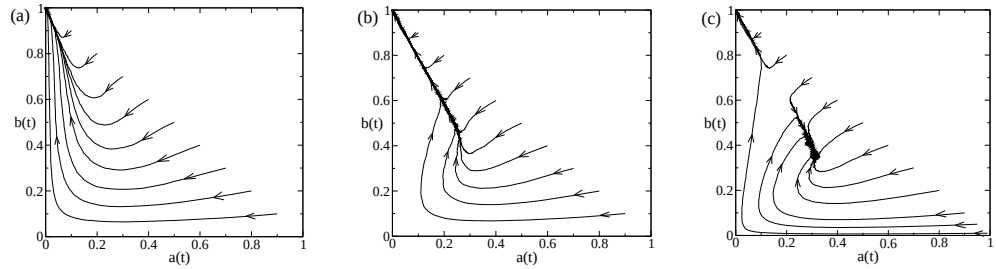


Fig. 2. Reproduced from [Mob1]. Effect of stirring on the phase portrait of the 2D stochastic lattice predator-prey with NNN interaction (rates $\eta = \mu = \sigma = 1$, $\delta = 10$, system size: 256×256). From left to right: 'stirring' rate $\mathcal{D} = 0, 2, 5$.

4. Mean-field behavior through short-range stirring. We have outlined the robust properties of most lattice predator-prey models. Among these characteristics, we have seen that deep in the coexistence phase the systems display complex clusters of activities at the interfaces of which the dynamics takes place. These features are not affected by the diffusion of either predators or prey. On the other hand, it is important to identify if there are ecological mechanisms able to render stochastic predator-prey models more tractable by means of deterministic equations. This requires to find processes allowing to efficiently stir the system by bringing the reactants within the interfaces of the clusters. Arguably, the simplest candidate is the process allowing any neighboring sites to exchange their content (a prey can avoid an incoming predator): $X + Y \to Y + X$, with $X, Y \in \{A, B, \oslash\}$. To test the efficiency of this ingredient, we have focused on a model whose mean-field (MF) behavior is markedly different from that of Eqs. (1,2). Actually, it is natural to split the Lotka-Volterra predation reaction by introducing two independent time scales and the new processes: (i) reproduction of the predators in the vicinity of a prey (*favorable environment*), according to the *next-nearest-neighbor* (NNN) reaction $A+\oslash+B \to A+A+B$ (with rate proportional to δ); and (ii) consumption of a prey by a neighboring predator, as $A+B \to \oslash+B$ (with rate proportional to η). The MF equations for this model, which can be viewed as comprising also a nearest-neighbor reproduction reaction ($A + \oslash \to A + A$) taking place on a much longer time scale, read [Mob1]

$$\dot{a}(t) = \delta a(t)b(t)[1 - a(t) - b(t)] - \mu a(t); \quad \dot{b}(t) = \sigma b(t)[1 - a(t) - b(t)] - \eta a(t)b(t). \quad (3)$$

In stark contrast to Eqs. (1,2), these equations admit *two stable fixed points* above a given threshold δ_c: one is (always) a node and corresponds to an absorbing steady state (system full of prey), while the other fixed point (either a node or a focus) is reactive (coexistence of prey, with a density $< 1/2$, and predators). Hence, according to MF theory, for $\delta > \delta_c$ this model undergoes a *first-order phase transition*. To check when these MF predictions actually apply to the stochastic lattice model, the latter has been numerically simulated in the presence of the exchange process (with stirring rate proportional to $\mathcal{D}$) [Mob1]. It turns out that there is a subtle interplay between the NNN interaction and the exchange process. As illustrated in Fig. 2, the latter can be summarized as follows: (a) For vanishing mixing ($\mathcal{D}$ small compared to the other rates), fluctuations have a drastic effect and invalidate the MF picture in dimensions $1 < d \leq 4$, where the system undergoes an active-

to-absorbing state transition belonging again to the DP universality class. (b) When one allows for short-range particle exchange ($\mathcal{D} > 0$), the phase portrait flows change dramatically (Fig. 2, center). (c) When the exchange process becomes sufficiently fast (e.g. $\mathcal{D} \approx \delta$) a reactive fixed point is also available (in any dimension), as demonstrated in Fig. 2 (right). In this case, the system undergoes a first-order phase transition as predicted by the MF theory and, for 'fast' stirring, the latter becomes very accurate. We have also found that the stable reactive fixed point is either a node or a focus. In the latter case, the coexistence phase is again characterized by moving activity fronts but, as the system is more mixed, these clusters appear less prominent than in Fig. 1 [Mob1].

5. Conclusion. We have outlined the robust features of a class of stochastic lattice models with Lotka-Volterra interactions and discussed how their deterministic (mean-field) descriptions are altered by spatial fluctuations and correlations. We have also shown that the rate equations can aptly describe the dynamics of a stochastic model in the presence of an efficient short-range exchange process. We have illustrated this point by considering a system with NNN interaction which exhibits either a first- or second-order phase transition, depending on the stirring rate.

References

[Ant] T. Antal and M. Droz, *Phase transitions and oscillations in a lattice prey-predator model*, Phys. Rev. E 63 (2001), 056119; M. Kowalik, A. Lipowski and A. L. Ferreira, *Oscillations and dynamics in a two-dimensional prey-predator system*, ibid. 66 (2002), 066107.

[Jans] H. K. Janssen, *On the non-equilibrium phase-transition in reaction-diffusion systems with an absorbing stationary state*, Z. Phys. B: Condens. Matt 42 (1981), 151; H. Hinrichsen, *Non-equilibrium critical phenomena and phase transitions into absorbing states*, Adv. Phys. 49 (2000), 815.

[Lot] A. J. Lotka, *Undamped oscillations derived from the law of mass action*, J. Am. Chem. Soc. 42 (1920), 1595; V. Volterra, *Leçons sur la théorie mathématique de la lutte pour la vie*, Gautiers-Villars, Paris, 1931.

[May] R. M. May, *Stability and Complexity in Model Ecosystems*, Princeton University Press, Princeton, 1973; J. Maynard Smith, *Models in Ecology*, Cambridge University Press, Cambridge, 1974; J. D. Murray, *Mathematical Biology*, Vols. I/II, 3rd ed., Springer-Verlag, New York, 2002; D. Neal, *Introduction to Population Biology*, Cambridge University Press, Cambridge, 2004.

[Mob1] M. Mobilia, I. T. Georgiev and U. C. Täuber, *Fluctuations and correlations in lattice models for predator-prey interaction*, Phys. Rev. E 73 (2006), 040903(R).

[Mob2] M. Mobilia, I. T. Georgiev and U. C. Täuber, *Phase transitions and spatio-temporal fluctuations in stochastic lattice Lotka-Volterra models*, J. Stat. Phys. 128 (2007), 447.

[Wash] M. J. Washenberger, M. Mobilia and U. C. Täuber, *Influence of local carrying capacity restrictions on stochastic predator-prey models*, J. Phys. Condens. Matter 19 (2007), 065139.

STOCHASTIC MODELS IN BIOLOGICAL SCIENCES
BANACH CENTER PUBLICATIONS, VOLUME 80
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WARSZAWA 2008

STOCHASTIC EFFECTS ON BIODIVERSITY IN CYCLIC COEVOLUTIONARY DYNAMICS

TOBIAS REICHENBACH, MAURO MOBILIA and ERWIN FREY

Arnold Sommerfeld Center for Theoretical Physics (ASC) and Center for NanoScience (CeNS)
Department of Physics, Ludwig-Maximilians-Universität München
Theresienstrasse 37, D-80333 München, Germany
E-mail: tobias.reichenbach@physik.lmu.de, mauro.mobilia@physik.lmu.de, frey@lmu.de

Abstract. Finite-size fluctuations arising in the dynamics of competing populations may have dramatic influence on their fate. As an example, in this article, we investigate a model of three species which dominate each other in a cyclic manner. Although the deterministic approach predicts (neutrally) stable coexistence of all species, for any finite population size, the intrinsic stochasticity unavoidably causes the eventual extinction of two of them.

1. Introduction. Cyclic dominance of species has e.g. been observed in lizard populations [Sin] as well as in bacterial communities [Kerr]. Here, three different morphisms of lizards, resp. three strands of bacteria, compete in a non-transitive way: each morphism/strand outperforms another but is in turn dominated by the third. The resulting cyclic coevolutionary dynamics may feature biodiversity [Kerr], which constitutes a central topic in modern ecology [May, Neal] and environmental policy [CNRS]. Much interest and effort is therefore devoted to the theoretical understanding of such systems. In the context of evolutionary game theory [Hof], the paradigmatic model for cyclic coevolution is the well-known rock-paper-scissors game, where each strategy beats another but is itself defeated by a third. The replicator dynamics deterministically describes the system's time-evolution, yielding oscillatory behavior. Similar results were already obtained in the pioneering work by Lotka [Lot] and Volterra [Vol], who described fish catches in the Adriatic through nonlinear differential equations, exhibiting oscillations as well. Despite their success and popularity, as a common shortcoming, these approaches ignore

2000 *Mathematics Subject Classification*: Primary 92D25; Secondary 82-06.

Key words and phrases: competing populations, stochastic and deterministic models.

M.M. gratefully acknowledges the support of the German Alexander von Humboldt Foundation through the Fellowship No. IV-SCZ/1119205 STP.

The paper is in final form and no version of it will be published elsewhere.

 © Instytut Matematyczny PAN, 2008

any form of internal noise arising from finite-size fluctuations, which are unavoidable for finite populations, as well as from the stochastic spatial distribution of the reactants.

In this article, we will specifically focus on the effects of finite-size fluctuations on the coevolution of a three-species model. Actually, the deterministic approaches (tacitly) assume that any finite-size effects should disappear in the limit of large populations, where they are expected to become valid. Only recently, the influence of the finite character of populations, as well as the relation between stochastic models and their deterministic counterparts have been investigated (see e.g. [Tra1, Tra2]).

In this article, we consider a paradigmatic microscopic model for cyclic coevolutionary dynamics and study the effects of internal stochasticity as compared to the deterministic predictions. We show that finite-size fluctuations dramatically alter the system's behavior: While the rate equations predict the existence of (neutrally) stable cycles, for any size, stochasticity leads the system to the extinction of two species. Technical details of our analysis can be found in [Rei1], here we focus on a more general and intuitive discussion.

2. The model. We consider three populations, labeled as A, B, and C, which mutually inhibit each other: A invades B, B outperforms C, and C dominates over A, closing the cycle. For the stochastic dynamics, we consider a finite number N of overall individuals in an "urn" [Fel], i.e. in a well-mixed environment. Two individuals may react with each other at certain rates $k_A,\ k_B$ and k_C according to the following reaction scheme:

$$A + B \xrightarrow{k_C} A + A, \quad B + C \xrightarrow{k_A} B + B, \quad C + A \xrightarrow{k_B} C + C. \tag{1}$$

Such "urn models" are closely related to the Moran process [Mor], which describes stochastic evolution of finite populations with constant fitness.

Note that since the reactions (1) conserve the total number N of individuals, one is left only with two degrees of freedom.

3. Deterministic analysis. Let us start with the deterministic approach, in which the mean-values of the densities of species $A,\ B,\ C$, denoted respectively $a,\ b$, and c, evolve according to the so-called rate equations, which are easily obtained and read:

$$\dot{a} = a(k_C b - k_B c), \quad \dot{b} = b(k_A c - k_C a), \quad \dot{c} = c(k_B a - k_A b). \tag{2}$$

Due to the conservation of the total density (for commodity, we set $a+b+c = 1$) only two of the three equations are independent. In addition, the quantity $K = a(t)^{k_A} b(t)^{k_B} c^{k_C}$ is conserved by Eqs. (2) [Rei1] and, as for mechanical systems with conserved energy, is responsible for cyclic orbits in the phase space. In Fig. 1 (a), the solutions of Eqs. (2) are shown in the two-dimensional phase portrait (simplex) for different initial conditions. The flows are cyclic trajectories around the neutrally stable reactive fixed point $(a^*, b^*, c^*) = (k_A, k_B, k_C)/(k_A + k_B + k_C)$, which corresponds to a coexistent state with nonvanishing densities of all species. On the contrary, the corners of the simplex are associated with absorbing states, where only one species survives. In addition, the boundary of the simplex is absorbing: once it has been reached, one of the species gets extinct and, among the two remaining populations, one dominates the other and the system is driven to one of

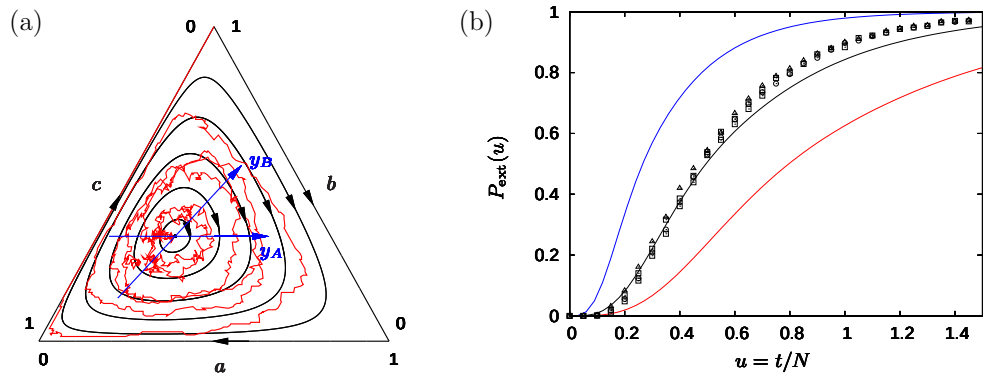


Fig. 1 (color online). Reproduced from [Rei1]. The three-species simplex for reaction rates $k_A = 1$, $k_B = 1.5$, $k_C = 2$ is drawn in (a). The rate equations predict cycles, which are shown in black. Their linearization around the reactive fixed point is solved in proper coordinates y_A, y_B (blue or dark gray). The red (or light gray) erratic flow is a single trajectory in a finite system ($N = 200$), obtained from stochastic simulations. It spirals out from the reactive fixed point, eventually reaching an absorbing state. The extinction probability $P_{\rm ext}$ is reported in (b) as function of the rescaled time $u = t/N$. Numerical data are shown for different system sizes $N = 100$ (triangles), 200 (boxes), 500 (circles). The red (light gray), blue (dark gray), and black curves correspond to analytical results for different values $R = 1/\sqrt{3}$ (top), $1/3$ (bottom), and their average value (middle).

the corners. As a consequence of the conservation of K,

$$\mathcal{R} = \mathcal{N}\sqrt{a^{*k_A} b^{*k_B} c^{*k_C} - K} \quad , \tag{3}$$

is also left invariant by Eqs. (2). Here, $\mathcal{N}$ is a normalization factor. As any perturbations to one closed orbit of Fig. 1 (a) give rise to a different regular cycle, the latter are said to be *neutrally stable*.

The quantity $\mathcal{R}$ monotonically increases with the distance from the steady state where it vanishes. Hence, it provides an efficient measure of the separation length between the reactive fixed point and any deterministic cycles. Furthermore, in the vicinity of the stationary state, the linearization of Eqs. (2) gives an accurate description of those orbits, which become circles in well suited coordinates (y_A, y_B in Fig. 1 (a)). In the linear regime, $\mathcal{R}$ is the radius of these circles whose characteristic frequency is $\omega_0 = \sqrt{\frac{k_A k_B k_C}{k_A+k_B+k_C}}$ [Rei1].

4. Stochastic behavior. The stochastic model does not conserve the quantity (3) and, as a consequence, the cyclic structure of the flows is lost. In fact, when one takes internal noise into account, the trajectory in the phase space can jump between the deterministic cycles. As an illustration of this fact, in Fig. 1 we have reported one such stochastic flow. Here, starting from the reactive fixed point, the trajectory soon departs from the latter, spirals around erratically and deviates from it with time. At some point, such a flow hits the absorbing boundary and ends up in one of the absorbing states. This is the generic scenario followed by the system. Indeed, due to the neutrally stable character of the deterministic cycles, the stochastic trajectory can perform a 'random walk' in the phase

portrait between them. Eventually, the absorbing boundary is reached and coexistence of all three species is lost. Hence, stochasticity arising from finite-size effects causes the extinction of two species. In the following, we outline an analytical approach allowing for a quantitative description of the above behavior, with a proper treatment of the finite-size fluctuations (see [Rei1] for a detailed discussion). For the sake of simplicity, we assume equal reaction rates ($k_A = k_B = k_C = 1$) in the basic reactions (1) that define the stochastic process. As usual, the dynamics of the latter is encoded in the related master equation. From the latter, it is fruitful to proceed with a Kramers-Moyal expansion (see e.g. [Gar]) which leads, up to the second term, to a Fokker-Planck equation (FPE). To make further progress, we consider a small noise approximation and linearize the FPE around the reactive fixed point [Rei1]. We also exploit the above-mentioned symmetry of the deterministic cycles in the y-variables, shown in Fig. 1, and introduce the polar coordinates $r,\ \phi$. In the latter, the FPE is recast in the neat form

$$\partial_t P(r,\phi,t) = -\omega_0 \partial_\phi P(r,\phi,t) + \frac{1}{12N}\left[\frac{1}{r^2}\partial_\phi^2 + \frac{1}{r}\partial_r + \partial_r^2\right] P(r,\phi,t). \tag{4}$$

The interpretation of this equation is simple and enlightening. The first term on the right-hand side of (4) accounts for the deterministic oscillatory dynamics with frequency ω_0. The second term, which is a Laplacian operator in polar coordinates and is proportional to $1/N$, encodes the finite-size fluctuations and ensures the (isotropic) diffusive character of the dynamics. It tells that fluctuations decrease with the system size, N.

Ignoring the absorbing character of the boundary for a moment, the FPE (4) allows us to compute the mean-square displacement $\langle r^2(t)\rangle$ from the reactive fixed point, when initially starting from there. In this situation, the initial probability distribution is a delta peak at the reactive steady state, which broadens in time, remaining a spherically symmetric gaussian distribution. Eventually, we obtain $\langle r^2(t)\rangle = t/(3N)$. Thus, larger systems require a longer waiting time to reach a given deviation from the reactive fixed point.

5. Extinction probability. Taking the absorbing nature of the boundary into account, we now investigate the extinction probability $P_{\text{ext}}(t)$, which is the probability at time t that the system, which was initially in its reactive steady state, has reached the absorbing boundary and two species become extinct. Note that an initially spherically symmetric probability distribution, as a delta peak in our case, is left spherically symmetric by the FPE (4). The dependence on ϕ in the latter equation therefore drops out, and only the second term remains, describing isotropic diffusion. Finding $P_{\text{ext}}(t)$ thus translates into a first-passage problem to a sphere. Note that, due to our linearization at the reactive fixed point, the triangular-shaped boundary is mapped onto a sphere, which is of course inaccurate. However, we are able to account for nonlinearities in a pragmatic manner, as shown below. The solution to the first-passage problem to a sphere is well-known and may e.g. be found in [Red]. We therefore directly obtain the Laplace transform of P_{ext}, which reads

$$\text{LT}\{P_{\text{ext}}(u)\} = \frac{1}{sI_0(R\sqrt{s/12})} \tag{5}$$

where $u = t/N$ is the rescaled time. As already found before, increasing the system size N only has the effect of protracting the time at which extinction occurs. In Fig. 1 (b) we compare results from stochastic simulations (based on the Gillespie algorithm) to our analytical approximations. For the simulations, different total populations sizes N were considered and, when rescaling time according to $u = t/N$, the data are seen to collapse on a universal curve, validating the scaling result. As for the analytical curves, nonlinear effects were dealt with by considering different distances R to the absorbing boundary. Small/large values of R over/under-estimate the numerical results. However, for an intermediate value of R, we obtain an excellent agreement.

6. Conclusions. We have presented a simple stochastic model for cyclic coevolutionary dynamics. Biodiversity in the form of coexistence of all three species emerges within the deterministic approach, which leads to neutrally stable oscillations. As the fluctuations drive the system to extinction of two species, stochasticity invalidates the rate equations predictions. We have quantified this behavior by deriving an appropriate Fokker-Planck equation, using linearization around the reactive steady state and by exploiting the polar symmetry of the system. Finally, we want to note that in spatially extended systems additional effects might strongly affect the survival of the species. As an example, in a three-level ecosystem (albeit without cyclic dominance) traveling wave structures have ben shown to enhance species survival [Bla]. Such topics are the subject of a forthcoming publication [Rei2].

References

[CNRS] Le journal du CNRS, N. 196 (2006)

[Bla] B. Blasius, A. Huppert, and L. Stone, *Complex dynamics and phase synchronization in spatially extended ecological systems*, 399 (1999), 354.

[Fel] W. Feller, *An Introduction to Probability Theory and its Application*, 3rd ed., Vol. 1, Wiley, New York, 1968.

[Gar] C. W. Gardiner, *Handbook of Stochastic Methods*, 1st ed., Springer, Berlin, 1983.

[Hof] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics*, Cambridge University Press, Cambridge, 1998.

[Kerr] B. Kerr, M. A. Riley, M. W. Feldman, and B. J. M. Bohannan, *Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors*, Nature 418 (2002), 171.

[Lot] A. J. Lotka, J. Amer. Chem. Soc. 42 (1920), 1595.

[May] R. M. May, *Stability and Complexity in Model Ecosystems*, Cambridge University Press, Cambridge, 1974.

[Mor] P. A. P. Moran, *Random processes in genetics*, Proc. Camb. Phil. Soc. 54 (1958), 60.

[Neal] D. Neal, *Introduction to Population Biology*, Cambridge University Press, Cambridge, 2004.

[Red] S. Redner, *A Guide to First-Passage Processes*, Cambridge University Press, Cambridge, 2001.

[Rei1] T. Reichenbach, M. Mobilia, and E. Frey, *Coexistence versus extinction in the stochastic cyclic Lotka-Volterra model*, Phys. Rev. E 74 (2006), 051907.

[Rei2] T. Reichenbach, M. Mobilia, and E. Frey, Nature 448 (2007), 1046; Phys. Rev. Lett. 99 (2007), 238105.

[Sin] B. Sinervo B. C. M. and Lively, *The rock-scissors-paper game and the evolution of alternative male strategies*, Nature 380 (1996), 240.

[Tra1] A. Traulsen, J. C. Claussen, and C. Hauert, *Coevolutionary Dynamics: From Finite to Infinite Populations*, Phys. Rev. Lett. 95 (2005), 238701.

[Tra2] A. Traulsen, J. C. Claussen, and C. Hauert, *Coevolutionary dynamics in large, but finite populations*, Phys. Rev. E 74 (2006), 011901.

[Vol] V. Volterra, Mem. Accad. Lincei 2 (1926), 31.

STOCHASTIC MODELS IN BIOLOGICAL SCIENCES
BANACH CENTER PUBLICATIONS, VOLUME 80
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WARSZAWA 2008

EFFECTS OF COMPETITION AND PREDATION IN A THREE SPECIES MODEL

JANUSZ SZWABIŃSKI, ANDRZEJ PĘKALSKI and KAMIL TROJAN

Institute of Theoretical Physics, University of Wrocław
M. Borna 9, 50-204 Wrocław, Poland
E-mail: szwabin@ift.uni.wroc.pl, apekal@ift.uni.wroc.pl, ktrojan@ift.uni.wroc.pl

Abstract. A model which consists of a predator and two prey species is presented. The prey compete for the same limited resource (food). The predator preys on both prey species but with different severity. We show that the coexistence of all three species is possible in a mean-field approach, whereas from Monte Carlo simulation it follows that the stochastic fluctuations drive one of the prey populations into extinction.

1. Introduction. One of the most important problems in ecology is to investigate the question of coexistence of species. Starting with the pioneering works of Lotka [1] and Volterra [12] there are numerous papers by either biologists, mathematicians or physicists studying the dynamics of predator-prey systems or systems with competition (see e.g. [7,1,2,5,10,8]). There are also some studies on three species food chains (e.g. [4]). However, to the best of our knowledge an exhaustive analysis of three species models with both predation and competition interactions is still missing. Understanding the interplay between these two mechanisms is of great practical interest, since it could suggest methods for manipulating populations in biological control.

In this paper we present a model of three populations living in the same ecosystem – two of them are prey and one is a predator population. The prey populations differ in one aspect only – one is preferred by the predators as food. Since it is known [3] that sometimes in biological models results obtained via mean-field type approach are even qualitatively different from those coming from agent-based simulations models, we use here therefore two methods – mean-field leading to differential equations and simulations based on treating each individual independently.

2000 *Mathematics Subject Classification*: Primary 92D25.
Key words and phrases: predator-prey model, Monte Carlo simulation.
The paper is in final form and no version of it will be published elsewhere.

 © Instytut Matematyczny PAN, 2008

2. Mean-field type model. The ecosystem that we model consists of a predator (*PR*) and two prey species (*P1* and *P2*). The prey species compete for the same limited resource (living space) but have unlimited access to food. Both prey species are food for the predators, but with different preference. The equations defining the dynamics of the model we study are:

$$\begin{aligned}\frac{\mathrm{d}N_1}{\mathrm{d}t} &= rN_1\left[1-\frac{N_1}{K}-b\frac{N_2}{K}-c_1N_3\right],\\ \frac{\mathrm{d}N_2}{\mathrm{d}t} &= rN_2\left[1-\frac{N_2}{K}-b\frac{N_1}{K}-c_2N_3\right],\\ \frac{\mathrm{d}N_3}{\mathrm{d}t} &= r_3N_3[c_1N_1+c_2N_2-d],\end{aligned} \tag{1}$$

where N_1, N_2 and N_3 are the numbers of elements of the *P1*, *P2* and *PR* populations, respectively. The parameters r and r_3 are the natural birth rates of the species, d is the mortality rate of the predators and K the carrying capacity. The constant b represents effects of competition between *P1* and *P2*. Predation rates are given by c_1 and c_2. Substituting the rescalings $x = N_1/K$, $y = N_2/K$, $z = N_3/K$, and $\tau = rt$ we get

$$\begin{aligned}\frac{\mathrm{d}x}{\mathrm{d}\tau} &= x[1-x-by-\alpha z],\\ \frac{\mathrm{d}y}{\mathrm{d}\tau} &= y[1-y-bx-\beta z],\\ \frac{\mathrm{d}z}{\mathrm{d}\tau} &= \gamma z[\alpha x+\beta y-d],\end{aligned} \tag{2}$$

where $\alpha = c_1K$, $\beta = c_2K$ and $\gamma = r_3/r$. The nullclines analysis of Eq. (2) leads to seven fixed points, the first of which is the trivial empty state $P_1 = (0,0,0)$. The next two points are states with only one prey species present: $P_2 = (1,0,0)$ and $P_3 = (0,1,0)$. In addition, there are three points with one species absent: $P_4 = (\frac{1}{b+1}, \frac{1}{b+1}, 0)$, $P_5 = (\frac{d}{\alpha}, 0, -\frac{d-\alpha}{\alpha^2})$ and $P_6 = (0, \frac{d}{\beta}, -\frac{d-\beta}{\beta^2})$. And finally, there is exactly one steady state P_7, where all three species exist:

$$\left(-\frac{(b\beta-\alpha)d+\beta(\alpha-\beta)}{\beta^2-2\alpha\beta b+\alpha^2}, \frac{(\beta-\alpha b)d+\alpha(\alpha-\beta)}{\beta^2-2\alpha\beta b+\alpha^2}, \frac{(b^2-1)d+(1-b)(\alpha+\beta)}{\beta^2-2\alpha\beta b+\alpha^2}\right). \tag{3}$$

The stability of all equilibria, as well as the dynamics of the system, is of course dependent on the values of the model parameters. To focus our attention we will consider the special case

$$\gamma = 0.50, \quad b = 0.20, \quad d = 0.90, \quad \alpha = 25 \tag{4}$$

and concentrate only on how the dynamics changes if the parameter β representing predation effects on the *P2* species varies. The values (4) are chosen to roughly match the oscillation period obtained from Monte Carlo simulation (discussed briefly in the next section). In this case the points P_1, P_2, P_3 and P_4 are unstable, no matter what the value of β is. From the local stability analysis [11] it follows that the point P_5 is stable for $\beta > \beta_1 = 25.7469$. Similarly, we can show that the equilibrium P_6 is stable if the condition $\beta < \beta_2 = 24.2523$ is satisfied. Furthermore, note that $x_7(\beta_1) = 0$ and $y_7(\beta_2) = 0$. Thus the interior point P_7 joins the fixed point P_6 or P_5 exactly at the transition value β_1

or β_2. It can be deduced from the above discussion that the coexistence is possible only in a narrow range $\beta \in (24.2523, 25.7469)$ around $\alpha = 25$. Indeed, one can check that for these values of β the interior point is a stable spiral. Outside that range the prey favored by the predators as food will be eliminated. In Fig. 1, an example evolution of the system in the coexistence case is shown.

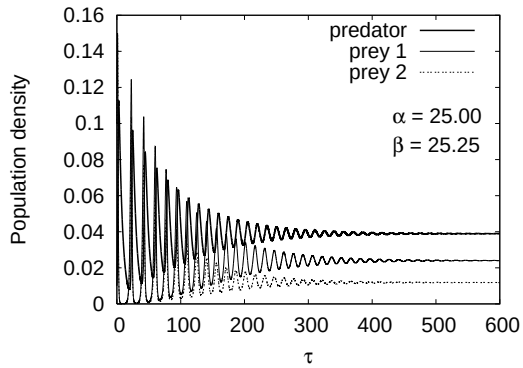


Fig. 1. Time evolution of the system in the coexistence case $\beta = 25.25$. The interior point P_7 is a stable spiral.

3. Agent-based model. Let us switch now to the Monte Carlo simulation of the investigated system. Our species live now on a square $L \times L$ lattice. They move at random and interact with each other following some rules. A detailed description of these rules may be found elsewhere [11], but we should mention some of them. (i) Double occupancy of sites by prey is not allowed. In this way the competition for living space is realized. (ii) Cross-breeding between two types of prey is prohibited. (iii) Predators reproduce if their food reserves are large enough. (iv) Animals reproduce in a quasi-sexual way, i.e. there is no sexual difference among the animals of the same type. However, two of them must meet in the neighboring sites in order to breed. (v) Prey is killed either with probability 1, or with probability $1 - p$, where the parameter p corresponds to the difference $|\alpha - \beta|$ in the mean-field aproach.

The model has the following parameters: (i) linear size L of the system, (ii) initial concentrations of the animals, (iii) birth rate m for prey, (iv) maximum food reserves f or alternatively, number of MCS a predator may live without catching prey, (v) minimum food reserves fm necessary for breeding, (vi) food reserves fb received by a baby predator and (vii) preference p in killing by predators of prey 1 over prey 2. It would be of course quite difficult, if not impossible, to investigate our model in the space of all those parameters. We have therefore decided to fix most of them and vary only L and p. The values taken for the remaining parameters are: initial concentrations $c(P1) = c(P2) = 0.2, c(PR) = 0.1, m = 6, f = 8, fm = 6, fb = 4$.

Our previous results show the existence of a narrow range of values of the p parameter, for which the coexistence of the three species is possible. To see if this is also the case in MC simulations, we let the system evolve for a given time and different values

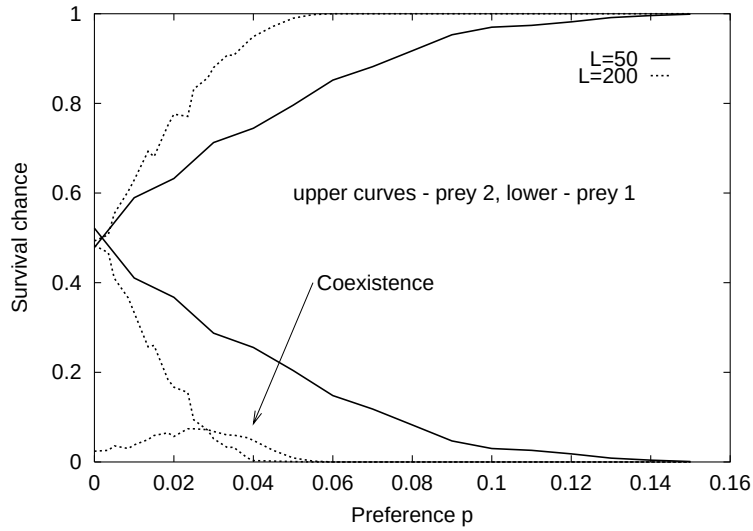


Fig. 2. Survival chance for the prey populations (at 6000 MCS)

of p and recorded the final densities of the species. The results (averaged over 2000 runs for $L = 50$ and 1000 runs for $L = 200$) after 6000 MCS are shown in Fig. 2. They are presented as a survival chance, i.e. the number of cases in which a given population was still present at the end of simulations. The difference between dynamics on larger ($L = 200$) and small (L=50) lattices should be noticed. For a relatively large value of p, say, $p = 0.05$, the chances of survival for *P1* in large lattices are nearly null. However on a small lattice they have a good chance to live on. The difference is due to stochastic fluctuations which play much more important role in small systems. This susceptibility of small populations to stochastic changes is well known in biology (see e.g. [9]). The role of the preference factor is, to some extent, specially in small systems, masked by stochastic effects in spatial arrangements of the population which may save or drive the population to extinction. The chance of coexistence is larger for bigger systems. For $L = 50$ we have never, regardless of the value of p, found a coexistence of the three species at the end of a simulation. For a larger system we have found a rather curious dependence of the coexistence chance on the value of p, since it initially increases with p.

Only when there are no predators, the two types of prey coexist forever and have the same concentration. In other cases there is a narrow range of the values of the preference parameter p, for which coexistence occurs, even for very long simulation times, however not indefinitely [11]. This kind of behavior differs from the mean-field results, but is quite similar to situations encountered in nature, where competition between two populations generally drives one of them to extinction if the living conditions are hard, but has small effect if the conditions are easy. This statement has never been proved in biology and is known as the Gause principle [13]. In our model difficult conditions are realized via the presence of predators.

References

[1] A. Arditi and L. R. Ginzburg, J. Theor. Biol. 139 (1989), 311.
[2] A. A. Barryman, Ecology 73 (1992), 1530.
[3] R. Durrett and S. Levin, Theor. Popul. Biol. 46 (1994), 363.
[4] A. Klebanoff and A. Hastings, J. Math. Biol. 32 (1994), 427
[5] A. Lipowski, Phys. Rev. E 60 (1999), 5179.
[6] A. J. Lotka, Proc. Natl. Acad. Sci. USA 6 (1920), 410.
[7] J. Maynard Smith and M. Slatkin, Ecology 54 (1972),384
[8] A. Pękalski, Computing in Science and Engineering 6 (2004), 62.
[9] M. Shaffer, BioScience 31 (1981), 131.
[10] D. Stauffer, A. Kunwar and D. Chowdhury, Physica A 352 (2005), 202.
[11] J. Szwabiński, A. Pękalski and K. Trojan, Int. J. Mod. Phys. C 17 (2006), 1629.
[12] V. Volterra, *Leçons sur la Théorie Mathématique de la Lutte pour la Vie*, Gauthier-Villars, Paris, 1931.
[13] M. Wiliamson, *The Analysis of Biological Populations*, E. Arnold, London, 1972.

STOCHASTIC MODELS IN BIOLOGICAL SCIENCES
BANACH CENTER PUBLICATIONS, VOLUME 80
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WARSZAWA 2008

A STOCHASTIC EXTENSION OF R. THOMAS REGULATORY NETWORK MODELLING

BARTEK WILCZYŃSKI

Institute of Mathematics, Polish Academy of Sciences
Śniadeckich 8, 00-956 Warszawa, Poland
E-mail: bartek@impan.gov.pl

Abstract. In this paper we present the extension of the kinetic logic proposed by René Thomas for analysis of genetic regulatory gene networks. We consider the case with a Gaussian noise added to the regulation function and propose a method of analyzing the resulting model with a discrete time Markov model.

1. Introduction. Understanding the regulatory mechanisms of gene expression is one of the key problems in molecular biology. Since such mechanisms are extremely hard to study *in vivo*, many mathematical models were proposed to help understanding the principles of regulatory network operation. The pioneering work in the field of regulatory network modelling was done in the 1960s by S. Kauffman [2] who showed that such fundamental phenomena of gene regulation as epigenesis and stable convergence can be modelled with a very simple mathematical framework of Boolean networks. This model was extended by Rene Thomas and co-workers [11, 12] leading to formulation of generalized logical description of regulatory networks.

Generalized logical modeling approach was applied to many experimentally studied biological regulatory circuits (e.g. [6, 10, 5, 4, 7]) showing that this formalism may be useful in regulatory network modelling. However, it is difficult to reconstruct such networks from experimental data. The problem with reconstruction of such networks lies in the lack of a natural scoring function for different models for a given dataset. Even though there are computational ways to effectively simulate such models (i.e. GINSim software package [3]), it doesn't help us with choosing the right model from many possibilities.

Another approach to this problem is using a stochastic dynamical system to model the dependencies of expression levels of genes. Chen et al. [1] managed to reconstruct

2000 *Mathematics Subject Classification*: Primary 92B99; Secondary 37H10.
Key words and phrases: gene networks, stochastic logical networks.
The author receives support from Foundation for Polish Science.
The paper is in final form and no version of it will be published elsewhere.

 © Instytut Matematyczny PAN, 2008

parameters for a system of stochastic differential equations from Yeast cell cycle expression data [9]. This was possible with the assumption that the expression of genes depends linearly on the expression of its regulators preventing the method to predict correct dependencies in cases where the regulation is non-linear.

In this work we present a novel modelling framework for regulatory networks, called Stochastic Logical Networks (referred to as SLNs). It is based on the formalism of generalized logical description of networks as introduced by Thomas [11] extended with the stochastic factor leading to a simple scoring function based on the likelihood of the model given the observed data.

2. Genetic network modelling

2.1. *Gene networks as dynamical systems.* In order to describe the formalism of SLNs, we provide a brief introduction into the way of modelling introduced by R. Thomas [11]. It is based on the assumption that a regulatory system can be accurately described as a dynamical system of ordinary differential equations. We treat the state of the cell, i.e. concentrations of all interesting gene products, as a vector of non-negative real values $\vec{v} = \langle v_1, v_2, \ldots, v_n \rangle \in R_+^n$, dependent on time t, so the equations have the following form:

$$\frac{\partial v_i}{\partial t} = -v_i \cdot \lambda_i + \mathcal{F}_i(\vec{v}), \tag{1}$$

where $\mathcal{F}_i(\vec{v})$ denotes the production rate of gene v_i depending on the state of all genes, whereas λ_i represents the decay constant responsible for degrading the gene product proportionally to its current concentration. To account for non-linearity and combinatorial nature of the dependence of the production rate on the state of regulators, the production rate of gene i is defined by Snoussi [8] as a linear combination of products of sigmoid functions of expression of regulators:

$$\mathcal{F}_i(\vec{v}) = \sum_{G \subseteq \{1 \ldots n\}} I_{G,i} \cdot \prod_{j \in G} S_{i,j}(v_j, \theta_{i,j}), \tag{2}$$

where $I_{G,i} \in \mathbb{R}^+$ is the regulatory influence of the set of regulators G on gene i, $\theta_{i,j} \in \mathbb{R}$ are the activation thresholds and $S_{i,j}$ is the sigmoid activation function of gene i by gene j being one of the following:

$$S^+(x, \theta) = sigm_k(x - \theta) \tag{3}$$

$$S^-(x, \theta) = 1 - sigm_k(x - \theta), \tag{4}$$

where $sigm_k(x) = (1 + e^{-kx})^{-1}$ with a notable case of $sigm_\infty$ equal to the Heaviside step function. Different forms of $S_{i,j}$ represent different possible regulatory interactions. If $S_{i,j} = S^-$ we say that j is a repressor of i, otherwise j is an enhancer of i.

2.2. *Qualitative approach.* René Thomas observed, that qualitative behaviour of such systems can be modelled as a non deterministic discrete process whose states correspond to discretized states of the original dynamical system. This is due to the fact that the production rates of genes change substantially only around the threshold values $\theta_{i,j}$. If we consider the case with $sigm_\infty$ step function, the hyperplanes $v_j = \theta_{i,j}$ divide the phase space of the dynamical system into a finite set of disjoint parts on which the production

rates of all genes are constant. In such case, the behavior of the system is determined by the choice of production rates of all genes in all discretized states. A simplistic example of a 2-gene negative feedback loop with its phase space, dependency graph and discrete state graph is depicted in Figure 1.

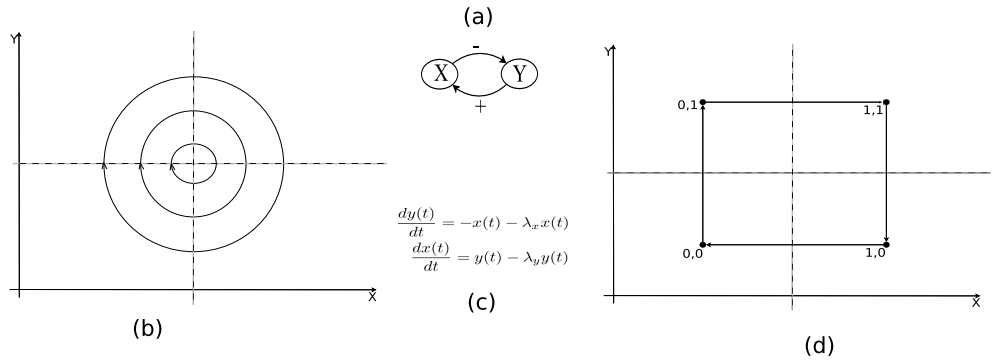


Fig. 1. An example of a dynamical system consisting of two genes: X and Y (a). Its phase space (b), ODEs (c) and state graph (d).

We use the notion of a discretization mapping $\delta(\vec{v}) = \langle \delta_1(v_1) \dots \delta_n \rangle$, where each δ_i is a mapping of the i-th variable into its discrete values using the thresholds $\{\theta_{i,j}\}_{j=1..n}$. We also denote the space of all discrete states by $\Sigma_n = \{0 \dots n\}^n$

Evolution of the qualitative system can be derived from equations governing the dynamical system. If we assume that $S_{i,j}$ are indeed Heaviside step functions, the value of the regulation functions $\mathcal{F}_i(\vec{v})$ is constant between the thresholds $\theta_{i,j}$. Therefore for each discrete state σ, corresponding to the discretization domain $\delta^{-1}(\sigma)$, there exist constant production rates $F_i(\sigma)$ of all genes, such that:

$$\forall_{\vec{v} \in \delta^{-1}(\sigma)} \mathcal{F}_i(\vec{v}) = F_i(\sigma).$$

However, the Heaviside regulatory functions introduce discontinuity in the right-hand side of ordinary differential equations for points in the state-space where $v_j = \theta_{i,j}$. We call such points *singular*, and exclude them from further analysis. This can be done without loss of generality since the measure of the set of singular points is 0.

After Thomas [11], we use the notion of an image function $\mathcal{R}$ of a discrete state σ:

$$\mathcal{R}(\sigma) = \langle \delta_1(F_1/\lambda_1), \dots, \delta_n(F_n/\lambda_n) \rangle.$$

If a state is the image of itself, we call it *stable*. Otherwise, since there may be different trajectories of the dynamical system traversing this discretization domain, the state succession is non-deterministic. For each discrete non-stable state σ we define a set of *successor states* $succ_\sigma$ containing all neighbouring discrete states σ' such that there exists a trajectory in the dynamical system going from $\delta^{-1}(\sigma)$ directly to $\delta^{-1}(\sigma')$. We use the notation of $\sigma \to \sigma'$ to denote the fact that $\sigma' \in succ_\sigma$. The generalization of a successor state is its transitive closure: the *reachability* relation $\sigma \to^+ \sigma'$.

2.3. *Stochastic logical networks.* We propose to introduce a stochastic factor to the differential equations describing regulatory systems. To this aim, we follow the common

practice of adding white noise in the form of an independent Wiener processes $W_i(t)$ scaled for each variable by ϵ_i to the right-hand side of differential equations (1) obtaining the following

$$dv_i = (\mathcal{F}_i(\vec{v}) - v_i \cdot \lambda_i)dt + \epsilon_i dW_i(t). \tag{5}$$

If we use the regulation functions $\mathcal{F}_i$ as defined in equation (2) and apply the same discretization mapping δ we obtain a stochastic system consisting of a finite set of disjoint domains. In each of these domains we observe a multivariate Brownian motion with linear drift – a very well studied mathematical model. What we are interested in, is a discrete stochastic process representing the movements of the Brownian motion between domains through time that correspond to the changes of qualitative behaviour of the whole system.

Because of the singularities in the right-hand side of stochastic equations at domain borders, we cannot analyze the exact dynamics of such systems in general. However, if we limit our analysis to the qualitative behaviour of the discrete system, we may make a simplifying assumption that the discretized process is Markovian. This is analogous to the definition of kinetic logic which assumes that successor states depend only on the current qualitative state of the system. It should be noted here, that at this point we trade the ability to analyze the subtle behaviour that may occur at the boundaries of domains for the simplicity of the discrete model.

For our considerations it is important to find the relationship between the parameters of the discrete process and the dynamical system. We need to answer the question: given a multidimensional Brownian motion with drift and a discretization of its state space into a set of disjoint domains what is the natural probability distribution of exiting from any of the domains to each of its neighbours. It is clear that the probability of Brownian motion exiting from any finite domain is equal to 1 and that the probability of moving directly to any non-adjacent domain is 0. We need to calculate the probabilities of travelling to all adjacent states for any given state.

Let us first examine the simple case of one dimensional Brownian motion with drift

$$\frac{\partial x}{\partial t} = F(x) + \epsilon \frac{\partial w}{\partial t}$$

and a finite closed interval $[a, b]$. We are interested in the probability $p_a(x_0)$ of exiting from $[a, b]$ through the barrier a starting from x_0. In general, we can determine this by solving the following differential equation

$$\frac{\partial^2 p_a}{\partial x^2} \cdot \frac{\epsilon^2}{2} + F(x)\frac{\partial p_a}{\partial x} = 0$$

assuming that $p_a(a) = 1$ and $p_a(b) = 0$. The solution of this equation depends heavily on F. In our case in each domain of continuity of the regulation function, $F(x) = C - x \cdot \lambda$. In this case we can obtain the following solution:

$$p_a(x) = \frac{\operatorname{erf}\left(\frac{\sqrt{2}\,(b\,\lambda - C)}{\epsilon^2\sqrt{-\lambda/\frac{\epsilon^2}{2}}}\right) - \operatorname{erf}\left(\frac{\sqrt{2}\,(\lambda\,x - C)}{\epsilon^2\sqrt{-\lambda/\frac{\epsilon^2}{2}}}\right)}{\operatorname{erf}\left(\frac{\sqrt{2}\,(b\,\lambda - C)}{\epsilon^2\sqrt{-\lambda/\frac{\epsilon^2}{2}}}\right) - \operatorname{erf}\left(\frac{\sqrt{2}\,(a\,\lambda - C)}{\epsilon^2\sqrt{-\lambda/\frac{\epsilon^2}{2}}}\right)}$$

where erf is the error function:

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt.$$

After simplifications we obtain the formula

$$p_a(x) = \frac{\int_x^b e^{-(\lambda(\frac{it}{\epsilon\sqrt{\lambda}} - C))^2} dt}{\int_a^b e^{-(\lambda(\frac{it}{\epsilon\sqrt{\lambda}} - C))^2} dt}.$$

Since the result is not elementary and we would like to have a reversible function we assume that the drift is constant and equal to c (it is reasonable in cases of small intervals). In this case our solution is

$$p_a(x) = \frac{e^{-\frac{2\,b\,c}{\epsilon^2}} - e^{-\frac{2\,x\,c}{\epsilon^2}}}{e^{-\frac{2\,b\,c}{\epsilon^2}} - e^{-\frac{2\,a\,c}{\epsilon^2}}}.$$

Now in order to have a solution not depending on the starting point, we integrate it over all possible starting points:

$$\int_a^b p_a(x)dx = \frac{\epsilon^2}{2c} - \frac{b-a}{e^{\frac{2c\cdot(b-a)}{\epsilon^2}} - 1}.$$

2.4. *Multidimensional case.* Now as we have the solution for one-dimensional case, we shall take multiple dimensions into account. As the stochastic components for different variables are independent, we can consider them as independent in each discretization domain.

Let us introduce, for each variable x_i, the probability $p_{i,a,b}(x,t)$ that the process starting from x will not leave the interval (a,b) in time t. What we want to know is, for each variable, what is the probability that it will be the first to leave the domain,

$$q_i(x) = \int_0^\infty \hat{p}_{i,a_i,b_i}(x,t) \prod_{j\neq i} p_{j,a_j,b_j}(x,t)dt,$$

where $\hat{p}_{i,a,b}(x,t)$ denotes the probability of the ith process reaching one of the points $\{a_i, b_i\}$ at time t starting from x.

In our case, in which the influence of drift is stronger than the influence of stochastic noise, $p_{i,a,b}(x,t)$ can be approximated by the probability $p'_{i,a,b}(x,t)$ of the ith process being in the interval (a,b) at time t. This can be computed from normal cumulative distribution function.

3. Conclusion. We present an approach extending the kinetic logic model of regulatory networks. We describe the continuous regulatory function as Langevin equations and under the simplifying assumption that the discrete process is Markovian, we are able to calculate the discrete state change probabilities using the Fokker-Planck equation. Even though such a simplified model is not suitable for quantitative analysis it may be useful in applications requiring more qualitative description. One such application is the task of network topology reconstruction from data. We have used this model for this task [13] with satisfactory results.

References

[1] K. Chen, T. Wang, H. Tseng, C. Huang, and C. Kao, *A stochastic differential equation model for quantifying transcriptional regulatory network in Saccharomyces cerevisiae*, Bioinformatics 21 (2005), 2883–90.

[2] S. Kauffman, *Homeostasis and differentiation in random genetic control networks*, Nature 224 (1969), 177–178.

[3] A. Larrinaga, A. Naldi, L. Sanchez, D. Thieffry, and C. Chaouiya, *GINsim: A software suite for the qualitative modelling, simulation and analysis of regulatory networks*, Biosystems 84 (2006), 91–100.

[4] L. Mendoza, D. Thieffry, and E. Alvarez-Buylla, *Genetic control of flower morphogenesis in arabidopsis thaliana: a logical analyssis*, Bioinformatics 7 (1999), 593–606.

[5] L. Sanchez and D. Thieffry, *A logical analysis of the Drosophila gap-gene system*, Journal of Theoretical Biology 211 (2001), 115–141.

[6] L. Sanchez and D. Thieffry, *Segmenting the fly embryo: a logical analysis of the pair-rule cross-regulatory module*, Journal of Theoretical Biology 224 (2003), 517–537.

[7] L. Sanchez, J. van Helden, and D. Thieffry, *Establishement of the dorso-ventral pattern during embryonic development of drosophila melanogasater: a logical analysis*, Journal of Theoretical Biology 189 (1997), 377–389.

[8] E. H. Snoussi, *Qualitative dynamics of piecewise-linear differential equations: a discrete mapping approach*, Dynamics and Stability of Systems 4 (1989), 189–207.

[9] P. T. Spellman, G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders, M. B. Eisen, P. O. Brown, D. Botstein, and B. Futcher, *Comprehensive identification of cell cycle regulated genes of the yeast saccharomyces cerevisiae by microarray hybridization*, Molecular Biology of the Cell 9 (1998), 3273–3297.

[10] D. Thieffry and L. Sanchez, *Dynamical modelling of pattern formation during embryonic development*, Current Opinions on Genetic Development 13 (2003), 326–330.

[11] R. Thomas, *Boolean formalization of genetic control circuits*, Journal of Theoretical Biology 42 (1973), 563.

[12] R. Thomas and R. D'Ari, *Biological Feedback*, CRC Press, 1990.

[13] B. Wilczyński and J. Tiuryn, *Regulatory network reconstruction using stochastic logical networks*, Lecture Notes in Computer Science 4210 (2006), 142–154. http://dx.doi.org/10.1007/11885191_10.